My Set-Top Box of Tomorrow is Not Your Set-Top Box of Today: Alternative to the VoD Services or Advances in the VoD?

Kyoungro Yoon*

1. INTRODUCTION

There have been a lot of research and efforts from academia and industries to make Video-on-Demand systems be the next generation TV services for the last couple of decades.[1] As a result, a few number of VoD systems have been delivered and are in service. However, compared to the hype of the VoD, it was not very successful in terms of the service availability. The major reason why it could not reach the expected point is due to the lack of capabilities in several fields that the VoD system depends on. Requirements on some devices such as storage devices and network bandwidths increase linearly with the increased number of the subscribers or the number of simultaneous users of the system.

With the advent of the digital TV services, which can give you high quality pictures of about 1920 × 1080 resolution, the requirements on the storage and network can get even more demanding than the requirements that we expected 5 year ago.

With todays broadband digital cable services and digital satellite broadcasting, we receive more programs than we can actually handle. In a typical digital broadcast environment, we receive more than hundred programs at a certain moment in time. In such an environment, a single smart set-top box can be a good alternative to the VoD service with the help of a smart agent. Todays smart set-top boxes called PDR(Personalized Digital Recorder) are armed with built in MPEG-2 encoder/decoder, high speed but low cost large capacity random access storage devices, and a smart agent. A class of set-top boxes will have intelligence of understanding internationally standardized meta-data such as MPEG-7[2] and/or TV-Anytime [3]. I envision that there will be a service provider sending out metadata streams conforming to MPEG-7 and/or TV-Anytime standard within the next 2-3 years. Popularization of such services and the standards can be another supportive technology of the VoD, since it can greatly simplify the process of locating desired content using an intelligent agent.

In this paper, I introduce the standards of MPEG-7 and TV-Anytime and exploit possibilities of the PDRs as an alternative to the

* LG Electronics Institute of Technology

VoD with MPEG-7 and TV-Anytime metadata service. Then, I will introduce the transitional PDRs which do not require the services of the MPEG-7 and/or TV-Anytime but can also be an alternative of the VoD service. I also investigate the impact of the metadata standards such as MPEG-7 and TV-Anytime on the VoD services. In conclusion, I compare the smart PDRs and the VoD service so that the application areas, which cannot be satisfied by the PDRs, can be identified.

2. MPEG-7

The so-called MPEG standards refer to all the standardization activities performed in the ISO/IEC JTC1/SC29 WG11. The people working at the ISO/IEC JTC1/SC29 WG11 or the working group itself are also called the Moving Picture Experts Group(MPEG). The MPEG working group has released several standards so far, including MPEG-1, MPEG-2, and MPEG-4. While these MPEG standards are developed to provide efficient way of compressing audio/visual data and synchronization of them, the new work items of the MPEG group, i.e., MPEG-7 and MPEG-21 are expanding the scope of the MPEG standard from the traditional compression standard to broader aspects.

The recent standardization efforts of the MPEG, MPEG-7, whose official name is the ISO/IEC IS 15938, is on the multimedia content description and is to be released on the September of 2001 as a standard. MPEG-7 is composed of 7 parts which are Systems, Description Definition Language, Visual, Audio, Multimedia Description Scheme, Reference Software, and Conformance. The part 1, systems, defines binary format of the MPEG-7 descriptions, access units of the binary description, and synchronization between the media and the description. The part 2, description definition language, defines the tools to define the MPEG-7 content description and to extend the MPEG-7 content descriptions for specific use. The tool defined in the part 2 is called the DDL in short and is based on the XML-schema[4] to provide extensibility, human readability, and ease of use. The part 3, visual, defines purely visual descriptions of the content such as color, texture, shape, and region descriptors. The part 4, audio, defines purely aural descriptions of the content such as silence, pitch, energy, tone and melody. The part 5, multimedia description scheme, defines the descriptions of the content which are neither purely visual nor purely aural, such as the structure of the content, concepts represented by the content, navigation and access methods, and description of the collections of the multimedia content. The part 6, reference software, provides software tools defined by the other parts of the standards. The part 7, conformance, defines the conformance. In other words, the part 7 contains the definition on what is conformant to the standard and describes the conformant ways of using the MPEG-7 descriptions defined in other parts.

These MPEG-7 based descriptions can be used in various applications such as the multimedia retrieval systems, multimedia database

systems, and multimedia filtering systems. Among the various applications of the MPEG-7, TV or broadcast applications are definitely major applications of the MPEG-7. Some of the descriptions of the MPEG-7, which can be used in the broadcast applications, include information related to the program guide such as genre, creator, actor, title, and publisher information and the user preference information. For example, a smart agent can create a personalized channel based on the content description information and the user preference information. Or an unattended smart agent can guess and record a certain number of programs based on a certain users preference information.

3. TV-Anytime

TV-Anytime(TVA) forum is another kind of standardization effort led by a group of people from the industry interested in providing personalized TV services such as TV broadcasters, internet broadcasters, content owners, equipment manufacturers, infrastructure providers, solution providers and related service providers. The objectives of the TV-Anytime forum include defining specifications that will enable applications for the consumer electronics platforms with local persistent storage. Defined specifications of the TV-Anytime should be network/delivery method independent and should support inter-operable and integrated systems from content creators to the content consumers. The developed TV-Anytime system should protect the all parties involved.

S-5 is to specify the mechanism and data to management and protect intellectual property rights related to the content and the metadata.TV-Anytime specification is composed of five parts, two of which are informative specifications and the rest are the normative specifications. First part, S-1, is an informative specification and on the phase 1 benchmark applications. The S-1 provides business models and functionalities that should be supported by the TV-Anytime systems. Based on the business models provided the TV-Anytime architecture can be verified and the support of the key business models can be ensured. Second part, S-2, is another informative specification and on the system description. S-2 gives description of the TV-Anytime system architecture example and provides walk through of the content referencing scenario and walk-through example of the TVA system. Third part, S-3, is the first normative specification in this series and on the metadata. In TV-Anytime, metadata refer to all the descriptive data about the content and are also called attractors as they can be used to attract consumers to the content. S-3 also defines user preference and usage history data as part of the metadata specification as they can be used to select or filter the content based on the attractors by agent software. S-3 specifies the format or representation of the metadata with the semantics and uses XML-schema language. The specified metadata includes program information, segment representation, navigation and access data representation and group of programs information as well as user description

data. Fourth part, S-4, is a normative specification and on the content referencing. The S-4 provides mechanisms for the location independent identification of content through the location resolution of the CRID(content reference identifier). Fifth part, S-5, is the last in the normative specification and on the right management and protection. Currently, specification of the S-5 is in progress to be finalized on June 2001, and the requirement specification of the right management(R-5) is released on December 2000. S-5 is to specify the mechanism and data to management and protect intellectual property rights related to the content and the metadata.

The Phase one of the TV-Anytime defines three types of business models, which are broadcast model, consumer response model, and bi-directional broadband model. The broadcast model assumes that there is no access to the service provider, i.e., unidirectional delivery environment. Therefore, the features of the TV-Anytime system of the broadcast model are limited to those which do not require return path to the service provider such as the use of the electronic program guide to capture the content, VCR like functionalities such as playback and rewind, highlight view, personal indexing, and consumer preference support. The consumer response model assumes the existence of the limited occasional return path. For example, a consumer platform with dial-up access to the server can be treated as such a kind of system. With limited return path support, the system can have features such as verification of the delivery and capture of the content, keeping updated listing and capture data, portable consumer profile and collectable usage data. The bi-directional broadband model assumes that the system is always on the network. Therefore, this model can enable additional features such as content delivery on-demand, remote storage of captured content and consumer controlled secure transfer of content.

4. PDR : The Smart Set-Tops

Personal Digital Recorder(PDR) is a consumer platform which can be used to record and playback audio-visual contents by receiving them either from the broadcast channels or through the network. The PDRs in this paper requires several key functionalities. The very basic of which is to have full-time access to the high bandwidth random access storage devices, either locally or through the network so that the viewers can record programs of their choice for later view. The second of which is the time-shift functionality, using which viewers can pause the live broadcast program and come back to the point where they left, at any time. The third of which is the support of multiple users. The PDR should be able to be personalized, hence can distinguish multiple users to load appropriate setting values for each identified user. The fourth is the support of the electronic program guide. Even in the uni-directional broadcast model, the PDR should receive the program guide regularly so that the users can select the desired content based on

it and the agent can perform various functions. The program guide should include available time of the content, title of the content, genre of the content and the access method. The available time of the content means the time when the content is available. In case of terrestrial TV broadcast, it is the broadcast time. The access method identifies how the set-top or the PDR can get the content. In other words, it identifies the channel of the broadcast in case of the terrestrial or cable TV broadcast, and whether further user authorization is required to get the content.

4.1. PDR without Service Subscription

PDR without service subscription is a PDR for the most of the todays TV viewers. The PDR can be a satellite TV receiver, a cable TV receiver, or a terrestrial TV receiver. The PDR should have an embedded functionality to extract features from the receiving programs and provide non-linear video browsing functionality. Most of the features introduced for the PDR of this class can also be employed in other classes of PDRs.

The Fig 1 shows a brief architecture of the PDR with non-linear browsing functionality. The solid lines denote the data flow, the dotted lines denote control flow, and the dotted box denotes the PDR non-linear video browsing related modules. Please note that the extracted feature is denoted by the dotted storage, as the extracted feature does not have to be stored in a persistent storage. The extracted feature need

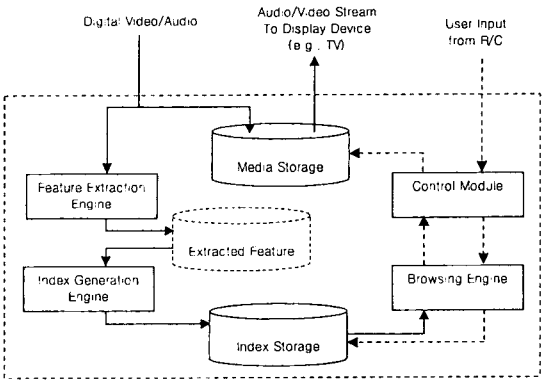

Fig 1. PDR with Non-Linear Video Browsing Capability

to be stored only for the duration of the index generation engine to process.

Researches on the multimedia information processing, specifically content-based multimedia information processing have been very active for the last couple of decades. However, there has not been any consumer platform available employing the results of the research on the content-based multimedia information processing. One of the main reasons is that most of the technologies developed so far very much depend on the characteristics of the target contents. While most of the content-based technology utilizes domain specific information, the current TV services do not provide any domain information such as the genre information. Therefore, only limited number of functionalities based on the multimedia information processing research can be applied to the consumer platforms, until the metadata service such as the EPG information service gets popular.

Some of the functionalities, which can be used in a consumer platform, are as follows:

1. Intelligent Fast-Forward/Reverse Play,
2. Structure-Based Instant Replay/Forward skipping,
3. Structure-Based Skimming,
4. And Thumbnail/Keyframe-Based Program List/Guide.

In the intelligent fast-forward/reverse play mode, the fast play speed changes depending on the perceptual complexity of the scene. For example, in the mid-field scene of a soccer game, the fast play speed can be maximized, as the users may feel that the scene is a very slow-going tedious scene. In a goal-area close-up scene of a soccer game, the fast play speed can be minimized, as the users may feel that the scene is a very fast-going exciting scene, or feel that it is too fast to understand the activities.

Structure-based instant replay/forward skipping provides an alternative to the conventional time-based instant replay/forward skipping. In the time-based instant replay/forward skipping, the skipping is strictly based on the time, e.g., seven seconds. However in the structure-based skipping, the skip command is replied by the play of the video program at the start of the current shot or the start of the next shot. In this way, the users need to check the missed portion of the content can be dynamically adapted based on the content change.

Structure-based skimming provides a primitive form of summarization of the selected content. In a regular skimming or fast-forward mode, there can be shots that are never presented or shots that are presented more than once as the segment selection is based on time. In the structure-based skimming, a portion of each shot is selected to be presented based on the shot-length. Therefore, each and every shot of the selected program is presented, helping the user understand the content of the selected program.

Thumbnail/Keyframe-based program list/guide can provide a user-friendly interface for selecting the recorded program. When each program is presented by the title, record-time, record-channel, and a thumbnail/keyframe, the thumbnail image can greatly help users to find the desired content.

When the PDR employs these features, there are various factors that should be considered as follows:

1. Real-time index generation
2. Minimum resource consumption
3. Simple and easy-to-use user interface.

When using a PDR, the users do not expect any latency when they request a feature such as the intelligent fast-forward or instant replay. Therefore, the feature extraction and index generation necessary to support the before mentioned functionality should be performed in real-time, so that the PDR is always ready for the user interaction.

While, the functionalities listed can be implemented using very complex algorithms to improve the accuracy, the trade-off between the accuracy and the resource consumption should be carefully considered. The central processor built into the PDR should perform various

activities such as user interaction interpretation, job management, as well as the feature extraction and index generation. Therefore, without consideration of the resource consumption, the price of the PDR can go unmarketably high. To keep the price of the PDR reasonably low, the resource consumption of the designed algorithms should be minimized as long as it provides reasonable accuracy.

The user interface should be designed very simple and easy to use. The PDR is not a general-purpose system such as a personal computer. Most of the users(including me) do not use complex features of the conventional VCR as most of the advanced features require multiple pushes of the buttons and are menu-based. For any good feature to give real benefits to users, they should be provided through very simple command such as a single or two pushes of the remote control buttons, or through a very intuitive user interface.

4.2. PDR as an alternative to the VoD service

With the advent of broadband digital cable and the digital satellite broadcasts, there are more channels and programs than a user can manually find and select. With addition of electronic program guide(EPG), which contains minimum information such as the time of broadcast, title of the program, genre of the program, and the channel information, the user can browse through the EPG and select the desired content. An advanced form of user interface can also enable users to search the desired content based on the EPG information. However, a naive user of the set-top box or the PDR would prefer pushing couple of buttons of the remote control device to typing in the keyword or the title of a movie.

A good consumer platform should provide very simple user interface so that even a naïve user can get the benefit of the technology or the convenience of the system. To satisfy such a requirement is the system resident smart agent. Based on the user preference such as preferred genre, title, and actors, the smart agent search through the EPG information and generates a list of programs of the users preference. When there is a enough space in the storage device, the agent can start recording a certain number of programs from the extracted preferable program list at the corresponding available time. Another way of using the smart agent is to create a virtual channel based on the users preference. The virtual channel can be presented in a chronological order with possibly a very few number of programs to select at a certain time. If the user wants, the agent start recording the content of a program while the user is watching another one. When the agent automatically records the candidate programs and a very large capacity storage device is prepared, the user can always select a program from the recorded programs. Such a feature of the smart-set top with EPG service can be a good alternative to the VoD service

The Fig 2 illustrates the process of building personalized channel based on the metadata

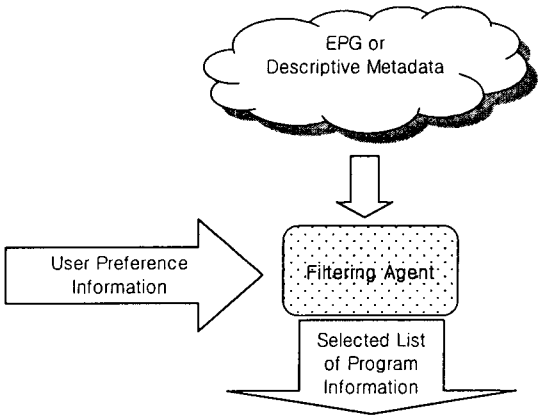

Fig 2. Building Personalized Channel Using Metadata and Preference Information

such as the MPEG-7 description, TV-Anytime metadata or Electronic Program guide information and the user preference information. The smart agent can take the metadata describing the available programs for a certain period time and generate a list of programs based on the registered users preference information. Once the list of programs is built, a virtual channel containing the selected programs from various channels can be created and presented to the users.

4.3. PDR as an advanced VoD client

On the other hand, a PDR with the interpretation capability of the metadata standards such as the MPEG-7 or TV-Anytime can greatly improve the service of the VoD. Fundamentally, the bi-directional broadband PDR of TV-Anytime is a VoD system, as it enables the users to request a specific content based on the metadata acquired. When a user is connected to the usual VoD service, it is very hard for the user to find out the exact program that he/she likes. Most of the VoD services provide a tree-structured list of programs and the users should navigate down the tree to see the available programs, or browse down the long list of programs. However, when the user does not know the exact title of the program, it is very hard for him/her to locate the desired program. With the rich metadata provided by the MPEG-7 and TV-Anytime, the users of the VoD can select the desired program based on various information that they know, e.g., actor name, director name, year of release, publisher, genre, and synopsis.

An example scenario of TV-Anytime PDR with bi-directional broadband network is as follows. The user has an access to the rich set of TVA compliant metadata describing available programs either by regularly receiving the metadata or by contacting the metadata server by network. The user selects one or more desired program by searching through the metadata. The PDR identifies the CRIDs of the selected program and sends them out to the service provider. The service provider locates the requested programs by using CRID resolution mechanism and delivers the programs to the requesting PDR. The PDR records or plays the incoming programs for the user abiding by the usage rules described in the right management metadata. A smart agent in this scenario can help the user to select the desired program with user-friendly interface or can even provide suggested list of programs based on the users personal preference information and the usage

history record.

As shown in the Fig 2 and the Fig 3, the basic concept of program filtering using the user preference information is identical with or without the VoD service, assuming the existence of the metadata. If the service provider provides MPEG-7 or TV-Anytime compliant description

5. Advances or Alternative?

Considering the capabilities of the various types of the PDR, PDRs can be divided into three types, which are the PDRs without service subscription, the PDRs with EPG service only, and the PDRs with metadata interpretation capability such as MPEG-7 or TV-Anytime, assuming all of those have non-linear video browsing features such as the time-shifting, intelligent fast forward/rewind, instant skip/replay, and skimming.

The PDRs without service subscription can

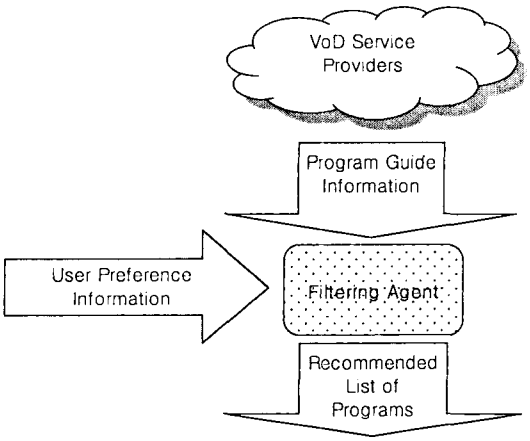

Fig 3. Using User Preference Information to Filter the Program List from VoD Service

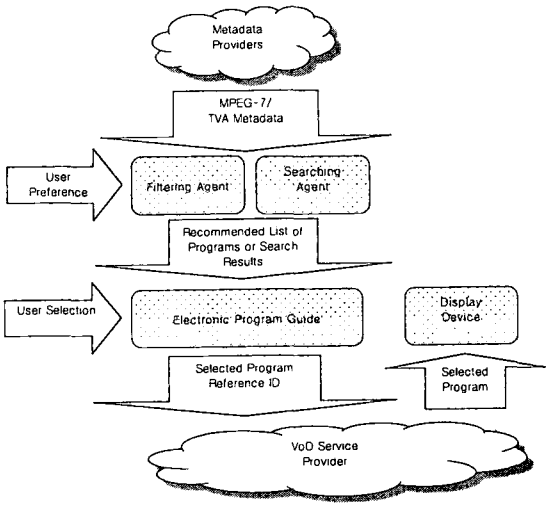

Fig 4. VoD Service with Metadata Service

provide very basic non-linear browsing features assuming the contents are recorded in the storage device. Unless there is any program guide channel which can be understood by the PDR, the capability of the PDR is very limited and cannot be considered as an alternative to the VoD service. However, with its non-linear browsing capabilities, it can be an advanced form of a VoD client.

The PDRs with EPG service can provide VoD-like service assuming there are enough number of channels and programs available. Based on the EPG information and user preference, the PDR can automatically capture the content without any user interaction. With its large size storage, it can provide a list of programs with the possible users interest. In such a sense, the PDR of this type can be a good candidate for the alternative to the VoD service. A VoD service usually provide menu-driven list of available programs, and EPG-like interface

of the stored programs can provide similar service. With their non-linear browsing capabilities and the local storage features, they can also be used as an advanced VoD client with added user control and back channel to the service provider.

The PDRs with MPEG-7 or TV-Anytime can provide very detailed search of the program based on the metadata compliant to the given standard. Once the content is identified, the PDRs can reserve to record the program when it is available or the PDR can play it when it is locally stored. Assuming the large capacity local storage device, the PDR can be an alternative to the VoD service with better selection capability than the conventional VoD service. The MEPG-7 or TV-Anytime standard can also provide better search and selection capability to the VoD service. Once the list of available programs is accompanied with the rich metadata specified by the MPEG-7 standard or TV-Anytime specification, the PDRs or the VoD client devices can provide a very detailed search capability resulting in a easier selection of the program.

As shown in this paper, the PDR can be used as both a VoD client and an alternative to the VoD. With the help of electronic program guide or MPEG-7/TV-Anytime compliant content description, the PDR with the metadata interpretation capability can be used as both VoD alternative and VoD client. The possibilities of the smart PDRs are summarized in the table [1].

Table 1. Comparison of Each Type of Set-Top/Service

	Advanced VoD Service	Alternative to VoD
PDR w/ Non-Linear Browsing only	+	-
PDR w/ EPG only	-	+
PDR w/ Metadata only	++	++
PDR w/ NLB & EPG	++	++
PDR w/ NLB & Metadata	+++	+++

In table 1, "-" means that the choice does not provide any significant improvement, and the number of "+" s means the significance of the improvement provided by the choice.

6. Conclusion

As we have studies in this paper, the next generation personal digital recorders(PDRs) compliant to MPEG-7 standard or TV-Anytime specifications can be greatly beneficial to the consumers as they can provide a functionality very similar to the video-on-demand(VoD) service without subscription to the VoD service. They can also provide very advanced features in terms of content filtering, selection, and viewing to the VoD service users as well as the non-linear video browsing functionality.

When the MPEG-7 standard or TV-Anytime based service is available, the consumption pattern of the TV programs will dramatically change. The viewers will spend less time looking for the programs of their interest, i.e., there will be less flipping of channels or scanning of the channels, and will spend more time watching their favorite programs. Also the viewers do not have to rush home to view their favorite programs, and enjoy the benefits of video-on-demand service with or without actual service, as they can view their favorite program

virtually any time they want.

In conclusion, the PDRs with the standardized metadata service and non-linear browsing features not only provide the most advanced features to the video-on-demand subscribers, but also provide very similar service to the consumers without the subscription to the video-on-demand service. A PDR with meta-data service and non-linear browsing features can be a good alternative to the video-on-demand service until the true VoD service is available by solving various problems of infra-structure including home network at the client side, and the broadband network and storage problem at the server side.

7. References

[1] Marek Pdgorny and Geoffrey C. Fox, Video on Demand Technologies and Demonstrations, Final Report, Northeast Parallel Architectures Center Contract No: F-30602-94-C-0256, Mar. 1997, available at http://trurl.webwisdom.com/rlvod/

[2] ISO/IEC FCD 15938-5 Information Technology Multimedia Content Description Interface Part 5 Multimedia Description Schemes, Mar. 2001, available at http://cselt.stet.it/mpeg/public/mpeg-7 mds fcd.zip

[3] TV-Anytime Forum Specification Series: S-3 on: Metadata (Normative), Document SP003V10, Feb. 2001, available at http://www.tv-anytime.org/

[4] XML-Schema Part 0: Primer, W3C Proposed Recommendation, Mar. 2001, available at http://www.w3.org/TR/xmlschema-0/

윤 경 로

- 1987년 3월 연세대학교 전자전산기 공학과 졸(학사)
- 1989년 12월 미시간대학 전기공학과 졸(석사)
- 1999년 5월 시라큐스대학 전산과학과 졸(박사)
- 1999년 6월 현재 LG 전자기술원 책임연구원
- 주관심분야: 멀티미디어인덱싱, 멀티미디어 정보처리