

CurioView: TV Recommendations Related to Content Being Viewed

Hideki Sumiyoshi, Masanori Sano, Jun Goto, Takahiro Mochizuki, Masaru Miyazaki, Mahito Fujii, Masahiro Shibata, and Nobuyuki Yagi

Abstract— We developed a new way of viewing TV, CurioView, which uses metadata and retrieval technology to satisfy viewers' curiosity by recommending wide-ranging video content related to the content the viewer is currently watching. We describe a general and expandable architecture that is based on CurioView's functions. The architecture can be applied flexibly, not just to TVs, but also to PCs and mobile terminals. We also report on the fundamental testing of a prototype system using this architecture.

Index Terms— IPTV & InternetTV, DTV and broadband multimedia systems, VoD, interactivity, datacasting

I. INTRODUCTION

THE rapid evolution of storage media and internet technology has made it easy to store large quantities of video content and distribute it across networks. But as more video content becomes available, it becomes harder for people to find things they want to watch. Search functions are essential for enabling large quantities of video content to be used more effectively, but it is unreasonable to expect television viewers to type in keywords or perform complex search operations. The role of television is changing due to the fusion of broadcasting and communication technologies. We think that television should have search and recommendation functions. Therefore, we have proposed a new viewing style called CurioView [1] that helps satisfy the diverse curiosity of television viewers by retrieving and recommending content related to what the viewer is currently watching. The content chosen by the viewer is assumed to reflect the viewer's interests at that point in time, so the metadata attached to this content can be used as a search key. This not only reduces the search operations that viewers have to perform, but should also be able to satisfy the curiosity of viewers in an intelligent manner by recommending related content.

We describe a model called CurioView, which is a flexible, general-purpose, and extensible content recommender system that not only works with television but can also be applied to other systems such as personal computers or mobile terminal. We also report on basic tests of a prototype system based on this model.

II. TRENDS IN CONTENT RECOMMENDATION TECHNOLOGY

There are several well-known techniques, such as GroupLens [2] and MovieLens [3], that generate recommendations through cooperative filtering that uses the previous actions and/or assessment results of other users. These are used on many e-commerce Internet sites, including Amazon's recommendation and the iTunes Genius feature, and are on their way to becoming standard features of systems that guide traffic to less-frequently accessed content or products. The Netflix DVD rental service held a successful contest to develop a better recommendation engine [4], which demonstrates that this approach is also effective for video content recommendations. However, in user environments where user's access is not enough coverage to the content, there will be more content outside the scope of the recommendation system. Under such circumstances, recommendations based on cooperative filtering may not be effective. Furthermore, since recommendations are made on the basis of a small number of actions or evaluations regardless of the nature of the content, there may be cases where incongruous products are offered as recommendations. Although such recommendations may be acceptable in some circumstances, they can sometimes leave users dissatisfied.

An alternative approach is a technique called content-based filtering, where knowledge about the subject matter is stored in a database and used to make recommendations. Researchers have applied this technique to the contents of research papers and online news articles [5][6]. In content-based filtering, all the products stored in the database are potential recommendations and can also be recommended to first-time users. Relationships between products can also be easily stored because the recommendations are based on the content details. However, it can be expensive to make and maintain the necessary database of content details and difficult to offer unexpected recommendations.

Work is currently underway on the development of hybrid techniques that combine content-based filtering and cooperative filtering [7][8]. Furthermore, methods have been proposed for assessing recommendations by focusing not only on the accuracy of recommendations but also on user satisfaction, such as the unexpectedness and acceptability of recommendations [9][10].

In CurioView, a content-based filtering technique is used because the primary objective is to make recommendations that

All authors are with Science & Technology Research Laboratories, Japan Broadcasting Corporation, 1-10-11, Kinuta, Setagaya-ku, Tokyo JAPAN (phone: +81-3-5494-3382; fax: +81-3-5494-3412; e-mail: sumiyoshi.h-di@nhk.or.jp).

satisfy viewers' interest in the content.

III. THE CURIOVIEW CONCEPT

We developed a new TV viewing style called CurioView, where TV programs (content) related to what the viewer is currently watching are recommended via an automatic search that uses metadata associated with the content. Instead of using fixed relationship data prepared by content providers, CurioView can use various search techniques to recommend related content that can be accessed at that point in time.

Although it uses search techniques, CurioView does not require viewers to perform complex operations such as entering search keywords or refining their search conditions as on a personal computer. They simply watch TV programs. It does this by regarding the content that a viewer is currently watching as a reflection of the viewer's interests at that point in time, and uses this content as a key to search for additional content while the viewer watches the program, as shown in Fig. 1. Recommendations are related to the current program and presented by using metadata associated with the content. Therefore, it should be possible to expand both the breadth and depth of the viewer's areas of interest as shown in Fig. 2 without requiring complex search operations.

In CurioView, a content-based filtering technique is used because the primary objective is to make recommendations that broaden interest in the content. In this technique, a database of metadata related to the content is used to make recommendations based on associations between content items, and any content can be selected as a recommendation.

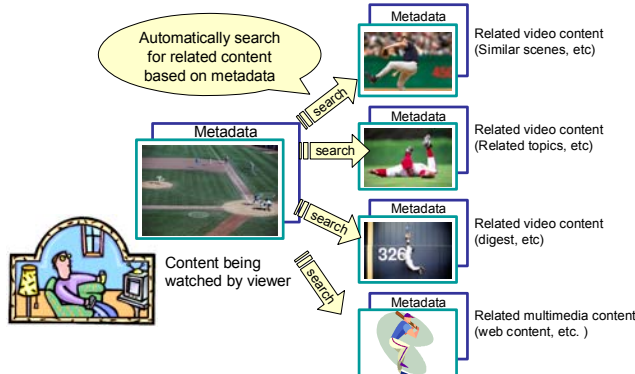


Fig. 1: The CurioView concept.

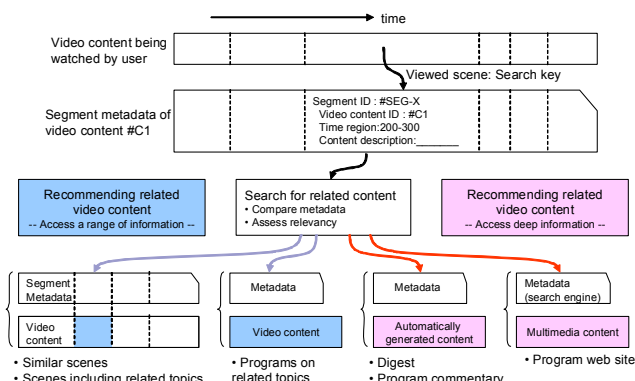


Fig. 2: The CurioView search mechanism and examples of content.

IV. CONFIGURATION AND FUNCTIONS OF CURIOVIEW

The CurioView system model and the conceptual interfaces between constituent elements are described below. The characteristic functions implemented are also described.

A. System model

CurioView is configured from the following four elements to make it easy to be functionally extended and used for a wide variety of different applications.

- (1) Related content display
- (2) Relevancy-based retrieval
- (3) Metadata management
- (4) Content delivery

To enable these four elements to operate cooperatively, we drew up the basic specifications of the functions of each element and the interfaces that connect them. The general flow of information and the operations of each part are shown in Fig. 3. The content delivery function can use diverse methods depending on how it is implemented (video format, etc.), so in CurioView the location of the content itself is confined to a logical functional definition where it is expressed in terms of metadata. This configuration can accommodate a wide variety of different display systems including TVs and personal computers or smart phones. It can also scale from small stand-alone systems to large-scale systems spread across multiple servers.

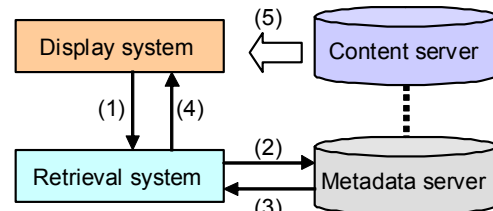


Fig. 3: CurioView system configuration.

The basic operations of the system's constituent elements are described below with reference to Fig. 3.

• Display system

The display system has content display and user interface functions. When the viewer watches content, the playback position (elapsed time from the start of the content) and an identifier (content ID) that uniquely identifies this content are sent to the retrieval system (1), which responds by sending back related content information to be displayed to the viewer (4). If necessary, the video may be displayed using the location information contained in the related content information (5).

• Retrieval system

The retrieval system provides a relevance-based retrieval function in response to requests from the display system. Metadata associated with the content that the viewer is watching is obtained from the metadata server (2)(3) by using the content ID and playback position information obtained from the display system (1). A query is then generated from this metadata, and related content is selected by a relevance evaluation program included in the retrieval system. The

content information (including the content title and storage location) on this selected content is sent back to the display system (4).

- Metadata server

The metadata server manages the metadata associated with the content. When it receives a request from the retrieval system (2), the metadata server responds by retrieving and sending back the metadata of the corresponding content (3).

- Content server

The content server delivers the requested content in response to the display system (5). The content storage location and access method are declared in the metadata.

B. Interfaces between elements

1) Retrieval interface

This interface is located between the display system and the retrieval system. The information included in requests issued by the display system is assumed to include items such as those shown in the sample below. These items can be adapted for a variety of different applications.

- Content ID
- Display time
- Rank and number of items returned

The retrieval system acquires information about the set of content obtained by query processing from the metadata server, and the following information is combined into a single content information element and sent back to the display system to consolidate the request.

- Content ID
- Associated label
- Similarity
- URL of actual content
- Title, subtitle
- Content summary text
- Genre
- Broadcast date and time

2) Metadata query interface

This interface is located between the retrieval system and the metadata server. The content ID is used as a query to acquire the metadata of the corresponding content. Management and a search of metadata become difficult when the quantity of content is large. We developed a metadata production framework (MPF) [11] for metadata production. Although we recommend that metadata be retrieved by using the query interface defined by the MPF, there may be CurioView implementations where metadata is managed integrally with the retrieval system, so the specific parameters are not defined.

C. Possible recommendation functions

With the system models and interfaces shown in sections A and B, it is possible to implement retrieval recommendation functions with a high degree of freedom. Examples include the following.

1) Recommendation based on content type or details

CurioView's system model can switch the recommendation function by type or by content details and put recommendation results together. On the basis of the genre and other content attribute information associated with the viewed content, suitable candidates can be selected from multiple relevance/similarity evaluation functions. Related content that is suited to the nature of the content being watched can be recommended to the viewer, which is enabled by embedded multiple relevance evaluation functions in the retrieval systems. Switching between retrieval processes using the metadata makes it unnecessary for the display system to specify the type or location of the retrieval processing. This increases the freedom of the retrieval system, makes it easier to extend its functions, and can reduce the implementation and processing time requirements of the display system.

2) Recommendation based on content granularity

The parameters of retrieval requests issued by the display system include the viewing content ID and current playback position as keys. In this way, the time information is able to acquire the metadata associated with a specified segment of the search source. This makes it possible to obtain search results that are tailored to the current playback position in the content (e.g., a particular news item or movie scene). The following four types of relevance retrieval are made possible by combining segment metadata with retrieval processing.

- whole content to whole content
- whole content to scene
- scene to whole content
- scene to scene

V. PROTOTYPE SYSTEM AND TESTING

We built a prototype system to demonstrate a specific example of the CurioView concept. As shown in Fig. 4, this system is configured on the premise that a network-connected television receiver or web browser is used as a display system for content stored, for example, as an archive or a video-on-demand (VoD) system. The retrieval systems are configured with the interfaces explained in section III.B implemented as HTTP SOAP WebServices.

The display systems and the retrieval systems at the core of the CurioView service are described below.

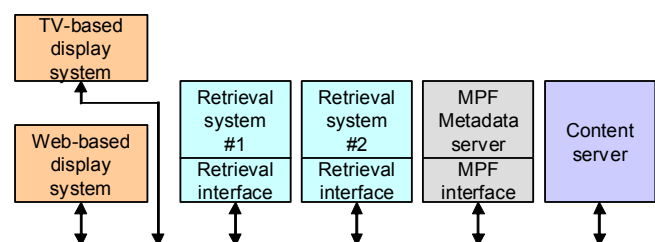


Fig. 4: Prototype system configuration.

A. Retrieval system

The retrieval system is implemented with the following three types of retrieval functions. These functions are selected on the

basis of the genre or the attribute information of the content being viewed.

- Using natural language processing to search for similar subjects

We have developed a technique for retrieving content on related subjects. This technique uses natural language processing to analyze electronic program guide text (or closed captions, speech recognition transcripts, program websites, etc.) related to the content, and calculates the similarity between the texts by focusing on the frequency with which characteristic words appear [12]. For news content, using the results of automatic speech recognition, the same method was used to extract related content by evaluating the relevancy of the text in each news item. We also implemented a technique for extracting related categories by focusing on terms that made a large contribution to the similarity (such as the names of people and words appearing in the topic).

- Using natural language processing to search the same event

We implemented a method for retrieving scenes of the same event from a video server that stored a baseball program by using natural language processing to analyze the newscaster's comments in news items about baseball games. This method of extracting the game progress and player names for each event discussed in the news program and comparing the results with the commercially available match metadata for each game that was manually entered by a specialist [13].

- Using video processing to search the same scene

We also implemented a method for retrieving a part of same baseball scene as a scene from a video server, which stores the baseball program containing the scene by using video processing to analyze the news video footage [14]. This method works by splitting the TV image into 12 regions and calculating four quantities—the average color “R”, “G”, “B” and the number of edge pixels (all normalized)—in each region to extract a vector with a total of 48 dimensions (4 quantities × 12 regions). By comparing the vector values at fixed frame intervals, we were able to extract scenes that most closely resembled the news footage from the live content stored on the server.

To make these three types of retrieval possible, we used metadata centered on the program summary text of approximately 4,000 TV programs, which was stored on a metadata server together with the video characteristic quantities (vector data) of several baseball broadcasts and the news footages.

B. Display system

The display system provides a user interface for issuing retrieval requests and displaying the search results. To check that the system model is capable of modifying the display modes and operations depending on factors such as the type of

user environment and whether or not a pointing device is available, we built one display system based on a TV receiver and another based on a web browser, and we verified the actions of both systems in combination with the same retrieval systems. We displayed related content recommendations using natural language processing to follow and retrieve similar topics. This was achieved with the following method.

1) Display system based on a TV receiver

For the TV-type display system, we opted to restrict the viewer operations to the use of a remote control unit, in consideration of the passive nature of TV viewing. From the usual full-screen viewing mode, the viewer can press a button to enter the CurioView mode where the display changes as shown in Fig. 5. A reduced view of the current program appears at the top left of the screen, while three content recommendations are shown on the right together with the related keywords that caused them to be recommended. The number of recommendations was provisionally set to three to facilitate their selection with a remote control unit, but we are currently investigating the ideal number of recommendations to offer. Also, since this prototype only displays a limited number of items, we decided to exclude series programs with the same name as the search source (like drama series, for example).

In the TV-type display system, considering the way people watch TV, we decided not to switch directly to viewing the recommended content. Instead we added a bookmark function to enable viewers to bookmark content for later viewing.



Fig. 5. Screenshot of CurioView running on a TV receiver

2) Display system based on a Web Browser

We also built a prototype display system using the Flash web browser plug-in. Unlike the TV display system, a computer user is able to operate a pointing device such as a mouse and is closer to the screen. We therefore adapted the system for use in a PC environment and made it capable of listing dozens of content recommendations as shown in Fig. 6.

The content list sent back from the retrieval system includes terms that indicate the degree of similarity and the type of association. This information is used to combine recommendations that have the same related term into a single group. Four groups of results are shown on the screen, comprising three clusters with the highest aggregate similarity and then recommendations combined into a fourth

group. Content with a high degree of similarity is placed close to the four corners of the video currently playing, which is shown at the center of the screen. The size of the thumbnail displays can be increased (in 5 increments) to represent their relevance to the content that is currently playing. The terms used to identify the relevance of this content are displayed at the four corners of the screen. User confidence can be increased by displaying a distribution of evaluations to indicate the relevance of or reasoning behind the recommendations [9]. Showing the terms that made a large contribution to the degree of similarity makes viewers more likely to accept the recommendation results. That research studied a collaborative filtering recommendation system, but we think it is applicable to content-based recommendation systems, too.

When the user hovers the mouse over a thumbnail displaying the recommended content, the content information is displayed in a text bubble, and the beginning of the content is played. Clicking the thumbnail selects the content. Content selected by the user is displayed in the central region, and then a new search based on that content is performed repeatedly.

With operations such as these, viewers can view multiple items of video content one after the other by following their own inclinations in much the same way as one might surf web pages on the Internet.



Fig. 6. Screenshot of CurioView running in a web browser

C. Basic tests

We demonstrated that TV-type and Web-type display systems could both be operated using the same retrieval systems. We were able to implement content recommendations that were successively adapted to dynamically changing content details by associating suitable segment metadata with different time slots of content such as news footage. We were also able to confirm that different retrieval processes such as natural language processing and image processing can be incorporated into similar frameworks, thereby verifying that this system model is highly extensible. Tests have demonstrated that the method used to select the retrieval process must be switched not by genre but in a way that

depends more on the nature of the content. Furthermore, tests have clarified problems with the implementation of systems that acquire information, such as the timing of retrieval requests.

VI. CONCLUSION

We described a new style of content recommendation for TV viewing called CurioView, and we have investigated a system model for implementing it as an extensible system. We built a small-scale prototype system and we confirmed its basic functions and extensibility (including its compatibility with multiple display systems and retrieval processes).

To produce content recommendations, a retrieval process for determining the relevancy of content is essential for judging what relationships exist between items of content. In our prototype system, similarity was taken as one criterion, but if the viewer's satisfaction is taken as an indicator, it is also important to present content that is related in unexpected tangential ways instead of just drilling down into content with the most relevance. Furthermore, an issue for future study is to verify and assess viewer satisfaction with the system's ease of use and quality of recommendations. We need to continue studying media processing techniques such as metadata extraction and retrieval techniques that extract various forms of relevance. We also need to investigate ways of making the system work faster (including a revised indexing scheme) so that video content retrieval and recommendations can be performed at practical speeds even with large-scale archives. We will also investigate mechanisms for making appropriate selections from among multiple retrieval process to suit the viewer's needs and for personalizing the recommendation results.

ACKNOWLEDGMENT

The TV-type CurioView display system was developed in cooperation with Panasonic Corporation, to whom we are very grateful.

REFERENCES

- [1] M. Fujii, M. Shibata, H. Sumiyoshi, J. Goto, M. Sano, T. Mochiduki, M. Miyazaki, N. Yagi, "CurioView: New viewing style of TV utilizing information retrieval -A study on embodying CurioView-", *Proc. ITE Annu. Symposium*, 6-5, 2009. (in Japanese)
- [2] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", *Proc. Conf. on Computer Supported Cooperative Work*, 1994, pp.175-186.
- [3] MovieLens, <http://www.movielens.org/>
- [4] Netflix Prize, <http://www.netflixprize.com/>
- [5] K. Lang, "NewsWeeder: Learning to Filter Netnews", *Proc. of the 12th Int. Conf. on Machine Learning*, 1995, pp.331-339.
- [6] S. Lawrence, C. Lee Giles, K. Bollacker, "Digital Libraries and Autonomous Citation Indexing", *IEEE Computer*, Vol.32, No. 6, pp. 67-71, 1999.
- [7] A. Das, M. Datar, A. Garg, "Google News Personalization: Scalable Online Collaborative Filtering", *Proc. 16th Int. Conf. on World Wide Web*, 2007, pp 271-280.
- [8] K. Yu, A. Schwaighofer, V. Tresp, W. Ma, H. Zhang, "Collaborative Ensemble Learning: Combining Collaborative and Content-Based Information Filtering via Hierarchical Bayes", *Proc Uncertainty in Artificial Intelligence*, vol.19, 2003, pp.616-623.

- [9] R. Sinha, K. Swearingen, "The Role of Transparency in Recommender Systems", *Proc. SIGCHI Conf. on Human Factors in Computer Systems*, 2002, pp. 830–831.
- [10] K. Swearingen, R. Sinha, "Beyond Algorithms: An HCI perspective on recommender systems", *ACM SIGIR Workshop on Recommender Systems*, 2001.
- [11] M. Sano, H. Sumiyoshi, M. Shibata, N. Yagi, "Metadata Production Framework (MPF) Version 2.0 - Designed for Effective Generation of Content-based Metadata", *Proc. ACM Int. Conf. Multimedia*, Oct. 2009, pp.1017–1018.
- [12] J. Goto, H. Sumiyoshi, M. Miyazaki, H. Tanaka, M. Shibata, A. Aizawa, "Relevant TV Program Retrieval using Broadcast Summaries", *Proc. ACM Int. Conf. on Intelligent User Interfaces*, Feb. 2010.
- [13] M. Miyazaki, H. Sumiyoshi, J. Goto, M. Fujii, M. Shibata, "Scene Recommendation System for Professional Baseball using Linguistic Information of Sports News", *Proc. Conf. on Forum on Information Technology*, No.2, F-008, 2009, pp.407–408. (in Japanese)
- [14] T. Mochiduki, M. Fujii, N. Yagi, K. Shinoda, "Automatic Event Classification in Baseball Broadcast Videos using Patternization of Scenes Focusing on Next Shot of Pitch and Discrete Hidden Markov Models", *The Journal of the Institute of Image Information and Television Engineers*, vol.61, no.8, 2007, p.1139–1149. (in Japanese)