Meaningful Scene Filtering for TV Terminals

Sung Ho Jin, *Student Member*, IEEE, Jun Ho Cho, and Yong Man Ro, *Senior Member*, IEEE

Abstract — *Nowadays, the progression of digital broadcasting has been toward establishing a user-friendly TV watching environment through personalized broadcasting services. In the consumer domain, video content filtering enables personalized broadcasting service in which a TV viewer obtains desired scenes even from live broadcasts which can not possess metadata. In this paper, therefore, we propose a new real-time content filtering system for live broadcasts which can be applied to TV terminals such as the set-top box and personal video recorder. Based on the proposed filtering, the TV viewer can watch not only favorite broadcasts of a main channel, but also get desired scenes from live broadcasts of other channels simultaneously. The usefulness of the proposed broadcasting content filtering is verified with sport videos on a test-bed broadcasting system*[1].

Index Terms —Real-Time Content Filtering, Personalized TV watching.

I. Introduction

With the appearance of digital broadcasting, the number of broadcasting channels has been increasing and TV terminals including set-top boxes (STB) and personal video recorders (PVR) have been distributed widely. In such an environment, broadcasting services have become personalized in recent years.

Due to the significant increase in the number of broadcasting channels, it is not easy for TV viewers to find their favorite programs quickly and efficiently. And, it has also become more difficult for viewers to obtain meaningful scenes from live broadcasts. The metadata does not support content-level indexing or retrieval in live broadcasting because it should be made before being put on the air.

Thus, a broadcasting system needs to provide a personalized broadcasting service scenario for live broadcasts. Let us suppose that a TV viewer is watching his favorite program (*e.g*., a drama or sitcom) on a main channel while other channels are broadcasting his second favorite program (*e.g*., the final round of a soccer game). The scenario is that he desires to get and watch the scenes of interest (*e.g*., shooting or goal scenes) on the second channel immediately while continuing to watch the main program.

[1] This work was supported in part by the Ministry of Information and Communication, Republic of Korea.

Sung Ho Jin is with the Image and Video System Lab, Information and Communications University (ICU), Daejeon, Republic of Korea (e-mail: wh966@icu.ac.kr).

Jun Ho Cho is with the Image and Video System Lab, ICU, Daejeon, Republic of Korea (e-mail: chojunho@icu.ac.kr).

Prof. Yong Man Ro is a director of the Image and Video System Lab, ICU, Daejeon, Republic of Korea (e-mail: yro@icu.ac.kr).

Current broadcasting systems provide simple program guiding services with electronic program guides that do not provide meaningful scene searching for live broadcasts [1-2]. To achieve the above scenario, real-time content filtering of live broadcasts should be embedded in TV terminals.

Thus far, many scene detection and content-indexing techniques have been applied to video archiving systems for video summarization, video segmentation, content management, and metadata authoring for broadcasts [3-6]. Most approaches have developed content understanding tools to create metadata or indexing information in the content creator or service provider.

However, in order to provide content-level service in live broadcasts, a meaningful scene search in real-time is required in the TV terminals. Some approaches have been performed in STBs or PVRs after recording entire broadcasts [7-9], but they cannot provide their results on the fly. N. Dimitrova et al. [10] studied video analysis algorithms and the architecture for abstracted video representation in consumer domain applications and developed a tool for commercial skipping by detecting black frames and changes in activity [11].

Compared with these works, our approach focuses on establishing a system which enables the indexing and analyzing of live broadcasts at the content-level.

In this paper, therefore, we propose a new real-time broadcasting content filtering system in a TV terminal which is acceptable in a multi-channel environment.

II. Real-Time Content Filtering

In this section, a TV terminal containing a broadcasting content filtering module is introduced and a content filtering algorithm, which is used for real-time processing in multi-channel inputs, is described.

A. Proposed Broadcasting Content Filtering System for TV Terminal

The architecture of a TV terminal with the real-time content filtering function is illustrated in Fig. 1. We expect that the terminal has computing power as a multimedia center or a server in the home [12], and suppose that it has more than two TV receiver parts which accept broadcasts from several channels simultaneously. One receiver part receives a broadcast from one main channel to be watched, and another part gets broadcasts from the selected channels to be filtered. This architecture can, therefore, support the preferences of TV viewers who are interested in both the selected channels and the main channel.

In Fig. 1, a broadcast selected from the main channel is conveyed to the display unit after passing the receiver and the decoder part. The frame sampler and filter are required for the

Manuscript received January 15, 2006 0098 3063/06/$20.00 © 2006 IEEE

proposed real-time content filtering. The decoder part feeds video frames of input broadcasts to the frame sampler with a regular sampling frequency and then the filter fetches the sampled frames for the sampler to recognize meaningful events depending on the video frames.

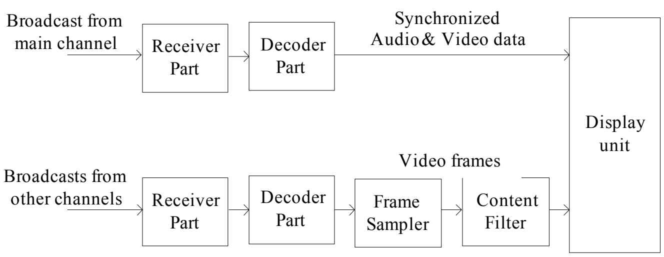


Fig. 1. Proposed architecture for a TV terminal containing TV watching and content filtering parts

B. Generic Real-Time Content Filtering Algorithm

The proposed filtering is based on the case where a service provider or a content creator cannot send metadata with broadcasts. Generally, news flashes or sport games are popular examples of this case. In particular, sport is an appropriate genre for filtering in TV terminals since many TV viewers are interested in sports videos and they consist of meaningful scenes that occur repeatedly.

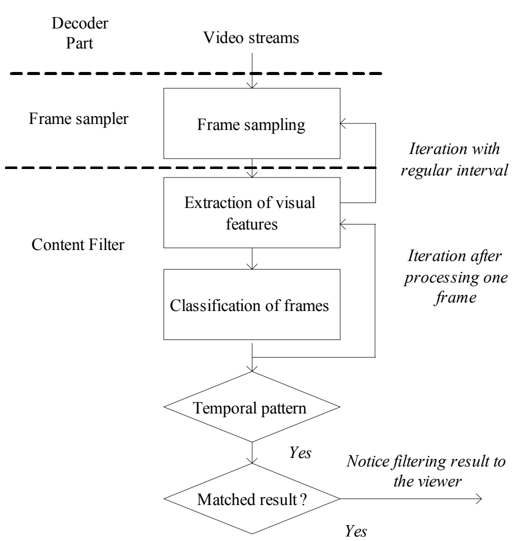


Fig. 2. Flowchart of the general filtering algorithm used in the proposed filtering system

Thus far, many previous works have developed effective event detection or scene extraction algorithms [13-17]. They show two approaches, with the first using *object-based* features (*e.g.*, object motions or trajectory). In sports video indexing, this approach enables high-level domain analysis, but extraction may be computationally costly for real-time implementation [13-14]. The second approach uses cinema features that result from common video composition and production rules (*e.g.*, replay), and threshold methods (as in the previous works) [15-17]. These minimize computational complexity but show limitations in high level indexing.

To achieve real-time processing, we employ a generic filtering algorithm using the cinema features and threshold methods of video frames as shown in Fig. 2.

The procedure of the content filtering process is described as follows.

Step 1: From the TV decoder part in Fig. 1, the frame sampler receives decoded video frames of broadcasts and samples them with a regular interval.

Step 2: Image and video features such as color, edge, and motion from the sampled frames are extracted to represent the spatial and temporal information of the frames. This step is repeated during the filtering process.

Step 3: The view type of the input frame is decided with visual features. After completing this step, a new frame is fetched for the next feature extraction.

Step 4: Desired scenes are detected by the pattern of temporal frame sequences such as shot detection, view type transition or continuity.

Step 5: Finally, the filtering process is concluded with the matched filtering result.

C. System Factors for Stability

Since the proposed content filtering should be performed within a given time, the system factors, which are defined as the frame sampling rate per channel, f, and the number of allowable input channels, N, should be suitably considered to achieve real-time processing.

For stable content filtering with a given terminal, the terminal selects the necessary factors satisfying the following conditions: 1) the filtering time (frame service time) maintains a faster speed than the sampling time (frame inter-arrival time); 2) the frame sampling rate guarantees minimal filtering performance.

In order to find the above factors, the proposed filtering system with N input channels is modeled as shown in Fig. 3.

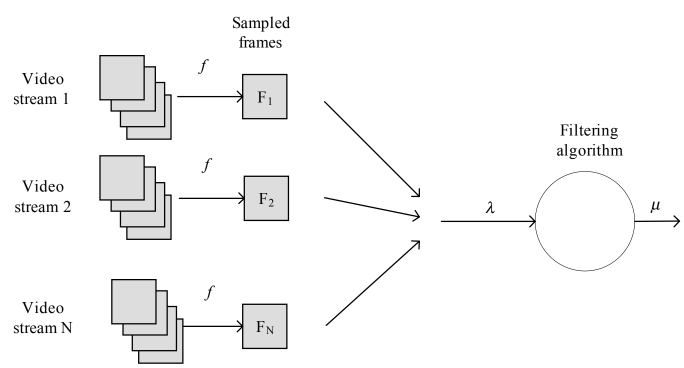


Fig. 3. Model of the proposed content filtering system with multiple channel inputs

In the proposed system, frames are acquired by sampling video streams and arrive in the filtering algorithm periodically. Let λ be denoted as the average arrival rate (the number of sampled frames per second), and μ be the average filtering rate which is the inverse of the average filtering time in the filtering process. Therefore, the arrival rate and the filtering rate can be written as

$$\lambda = N \cdot f \quad (1)$$

and

$$\mu = \frac{1}{\max_{x \in V}\{PT(x)\}} \quad (2)$$

, respectively.

Where, V denotes a set of frame view types, $\{v_1, v_2, \ldots, v_n\}$, for the input video, and $PT(x)$ represents the filtering processing time for x frame type. To find the sufficient value for the average filtering time, we can suppose that it is equal to the maximal value among filtering times for various frames.

Let ρ be denoted as the traffic intensity of the model which is equal to the arrival rate over the service rate (λ/μ). Considering that the proposed model is always stable if $\rho \leq 1$, the first condition can be established as follows,

$$\rho = N \cdot f \cdot \max_{x \in V}\{PT(x)\} \leq 1 . \quad (3)$$

The second condition can be established from the trade-off between frame sampling rate and filtering performance. The filtering performance, FP, depends on the sampling rate, f. It decreases as the frame sampling rate decreases. The relationship between them is denoted by the function of $FP(f)$. Thus, an appropriate sampling rate should satisfy the confidence limit of the filtering performance. The confidence limit, T_{fp}, denotes the minimum filtering performance value which TV viewers accept; namely, the filtering system should maintain this performance at least during the change of the sampling rate or the number of channels. This leads to the second condition which is expressed as

$$f > f_s , \quad (4)$$

where $f_s = \arg\min_f \{FP(f) \geq T_{fp}\}$.

If these two conditions are satisfied, the proposed filtering system can guarantee stable filtering in live broadcasts. We apply these two conditions to the content filtering system for soccer videos in order to verify the effectiveness of the model.

III. Implementation

In previous works, an intelligent broadcasting system was designed and implemented using agent platforms to provide personalized broadcasting services [18]. The proposed content filtering was verified on the system.

A. Test-bed Broadcasting System

Agent platforms in the broadcasting system can provide a user-friendly watching environment by minimizing user operations and adding intellectual functions. For that reason, we designed and implemented an intelligent broadcasting system for personalized content consumption.

For intelligence and autonomy of the broadcasting system, we utilized FIPA-OS, one of the most popular agent platforms. The system is composed of three independent agent platforms: the User Terminal, which directly interacts with TV viewers; the Service Provider, which responds to a viewer's query and provides metadata or contents to User Terminals; and the Service Information Provider which is an intermediary system between User Terminals and Service Providers for exchanging data or necessary information [18]. Our previous broadcasting system, consisting of three agent platforms, appears in Fig. 4.

Fig. 4. An intelligent broadcasting system applied with agent platforms: (a) User Terminal, (b) Service Information Provider, and (c) Service Provider.

In this system, the proposed content filtering is established as an intelligent broadcasting function provided by the User Terminal.

B. Applied Filtering Algorithm

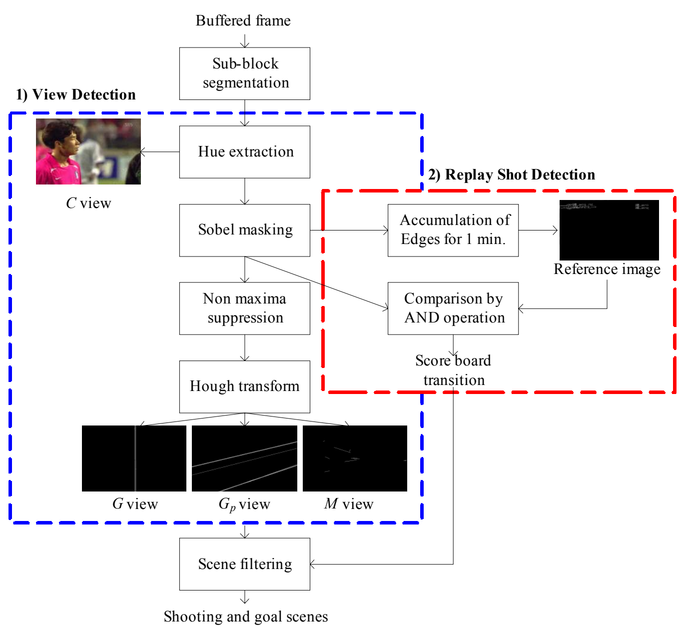


Fig. 5. Flowchart of applied filtering algorithm for soccer video

We apply the proposed real-time content filtering to soccer videos. First, we need to analyze soccer videos in order to extract meaningful scenes. In this paper, shooting scenes including goal scenes are used as the most important and favored scenes in soccer videos. The view types of frames in the video can be categorized into four classes: 1) a global view with goal post (G_p), 2) a global view without goal post (G), 3) a medium view (M), and 4) a close-up view (C), that is, $V=\{G_p, G, M, C\}$.

The flowchart in Fig. 5 shows that color features in the Hue domain and edge features are utilized to detect the view types. In order to minimize the detection time of the shooting scenes, a modified algorithm of [15] is used without shot segmentation. After identifying the temporal patterns by view detection, some verification time is inevitable for replay shot detection which verifies the filtering results of the meaningful scenes. Therefore, viewers obtain the filtering results with only a slight time difference.

C. Application with soccer video

Figure 6 represents the implemented real-time filtering of broadcasting content semantics on a soccer video. The main screen shows TV viewer's favorite programs on a music channel. And the dotted box on the bottom right-hand side of the screen in the figure broadcasts the desired broadcasting program from the channel of interest.

(a)

(b)

Fig. 6. Screen shot to run the real-time content filtering service with a single channel of interest: (a) The dotted box on the bottom right-hand side of the screen simulates the selected channel of interest (in this case, a soccer game); and (b) A TV terminal performs a channel change to give the result of the filtering.

In Fig. 6 (a), a viewer selects a soccer game on his/her channel of interest, and chooses "Shooting" and "Goal" scenes as the desired semantic information. On filtering them, the terminal interrupts the main screen by displaying the semantic video segment of the selected channel of interest as shown in Fig. 6 (b).

An agent existing in the User Terminal watches the channel of interest, which is selected by a TV viewer, instead of him/her. As soon as the agent captures his/her desired event by using the applied filtering algorithm, the terminal notifies the viewer of the fact by means of a channel change or channel interrupt function.

IV. EXPERIMENT

Experiments are performed with the soccer video database which consists of five soccer videos of about 90 minutes each in length.

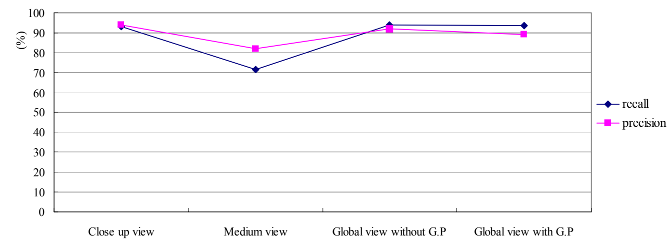


Fig. 7. Performance of the proposed view type decision algorithm

The filtering system is verified by evaluating the performance of the filtering algorithm including view decision and shooting scene detection in the soccer video. The performance of the used view decision algorithm is shown in Fig. 7. The experiments show that the total average of the recall rate is over 93.5% and that of the precision rate is about 89.5%.

When the proposed filtering algorithm is employed in an intelligent broadcasting system, the filtering performance of shooting scenes (including goals scenes) shows an average recall rate of 81.5% and an average precision rate of 76.4% in the database with 3 frames/second and one channel.

TABLE I
CONSUMING TIME FOR DECIDING VIEW TYPES

Terminal / View	T_1	T_2	T_3
Close-up	0.170 sec.	0.037 sec.	0.025 sec.
Medium	0.270 sec.	0.075 sec.	0.045 sec.
Global without G.P.	0.281 sec.	0.174 sec.	0.088 sec.
Global with G.P.	0.340 sec.	0.206 sec.	0.110 sec.

T_1 - 650 MHz CPU, 128MByte memory, T_2 - 1.7 GHz CPU, 512MByte memory, T_3 - 3 GHz CPU, 2GByte memory.

To find the factors satisfying the conditions for system stability, we measured the filtering times depending on the specifications of user terminals. The test-bed was set up with three terminals of different performances (T_1, T_2, T_3). The TV terminal plays a video on a channel and filters another video on another channel as well. In particular, a computer with a 650 MHz CPU is utilized to build similar capacity with the STB currently being used.

Considering that a major portion of the consuming time is to detect view types of input frames in the filtering system, we find how many channels are feasible in each terminal as the channel of interest. Table I shows the experimental results measuring the consuming times for deciding the view types per terminal.

It is shown that a global with a goal post, G_p, consumes the longest time. Then, substituting x into G_p in Eq. (3), the relationship between the number of channels, N, and the frame sampling rate, f, is given in Fig. 8. We can use the number of channels and the frame sampling rate which are placed under each curved line.

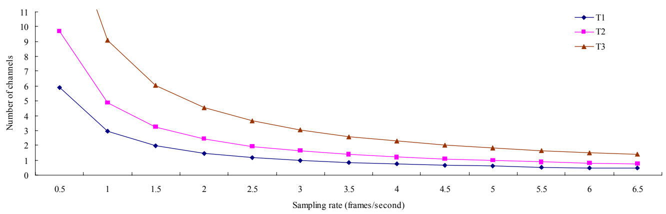


Fig. 8. Number of channels allowed by variation of sampling rate: T_1 has the lowest system capacity (650 MHz CPU, 128MByte memory). T_2 and T_3 own the second (1.7 GHz, 512MByte) and the highest system capacities (3 GHz, 2GByte), separately.

For the condition of Eq. (4), it is necessary to investigate the variation of filtering performance in terms of the change of sampling rate. Figure 9 shows that the minimum value of the frame sampling rate, f_s is determined by viewers' confidence limit, T_{fp}. In the figure, the solid line and dotted line are defined as the experimental results of filtering performance and the expected filtering performances, respectively. The expected values are acquired by removing the peaks of the experimental values to create a verified filtering performance. These values determine the minimum frame sampling rate.

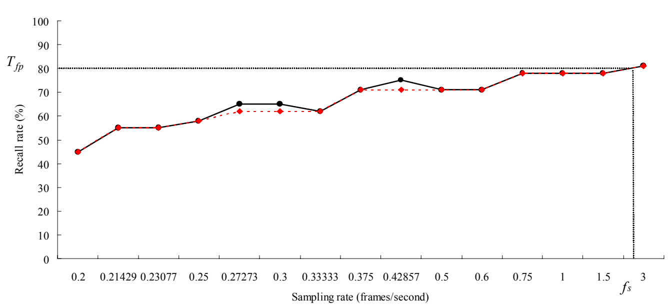


Fig. 9. Filtering performance depending on frame sampling rate. The solid line and dotted line are defined as the experimental results of filtering performance and expected filtering performance, respectively. From the dotted line, the system decides the minimum frame sampling rate, f_s.

From both Fig. 8 and 9, we can obtain the factors for the real-time content filtering system. Therefore, we implemented a filtering system, with 80% confidence limit and 3 frames per second, which guarantees stability conditions in the three terminals.

V. Conclusion

In this paper, we proposed a new real-time semantic filtering system to provide personalized and preferred scenes in broadcasts put on the air. From the fact that the experiments showed that even a low-performance PC with a 650 MHz CPU could perform the filtering function in real-time, we can draw an interesting conclusion which is that real-time content filtering in a TV terminal is possible with currently available STBs. Therefore, the proposed system and filtering method for real-time content filtering can provide TV viewers with a personalized broadcasting service. In terms of future work, more studies are needed to apply the proposed filtering to other genres and to analyze optimal system requirements.

References

[1] J. S. Choi, J. W. Kim, D. S. Han, J. Y. Nam, and Y. H. Ha, "Design and implementation of DVB-T receiver system for digital TV," *IEEE Trans. Consumer Electronics*, vol. 50, no. 4, pp. 991-998, 2004.

[2] M. Bais, J. Cosmas, C. Dosch, A. Engelsberg, A. Erk, P. S. Hansen, P. Healey, G. K. Klungsoeyr, R. Mies, J.-R. Ohm, Y. Paker, A. Pearmain, L. Pedersen, A. Sandvand, R. Schafer, P. Schoonjans, and P. Stammnitz, "Customized television: standards compliant advanced digital television," *IEEE Trans. Broadcasting*, vol. 48, no. 2, pp. 151-158, 2002.

[3] N. Dimitrova, H.-J. Zhang, B. Shahraray, I. Sezan, T. Huang, and A. Zakhor, "Applications of Video-Content Analysis and Retrieval," *IEEE Multimedia*, vol. 9, no. 3, pp. 42-55, 2002.

[4] X. Zhu, J. Fan, A. K. Elmagarmid, and X. Wu, "Hierarchical video content description and summarization using unified semantic and visual similarity," *Multimedia Systems*, vol. 9, pp. 31–53, 2003.

[5] C.-W. Ngo, Y.-F. Ma, and H.-J. Zhang, "Video Summarization and Scene Detection by Graph Modeling," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 2, pp. 296-305, 2005.

[6] H. Li, G. Liu, Z. Zhang, and Y. Li, "Adaptive Scene-Detection Algorithm for VBR Video Stream," *IEEE Trans. Multimedia*, vol. 6, no. 4, pp. 624-633, 2004.

[7] I. Otsuka, K. Nakane, A. Divakaran, K. Hatanaka and M. Ogawa, "A Highlight Scene Detection and Video Summarization System using Audio Feature for a Personal Video Recorder," *IEEE Trans. Consumer Electronics*, vol. 51, no. 1, pp. 112-116, 2005.

[8] J. Lee, G. Lee, and W. Kim, "Automatic Video Summarization Tool using MPEG-7 descriptors for Personal Video Recorder," *IEEE Trans. Consumer Electronics*, vol.49, no. 3, pp. 742-749, 2003.

[9] J. Kim, S. Suh, and S. Sull, "Fast Scene Change Detection for Personal Video Recorder," *IEEE Trans. Consumer Electronics*, vol. 49, no. 3, pp. 683-688, 2003.

[10] N. Dimitrova, T. McGee, H. Elenbaas, and J. Martino, "Video content management in consumer devices," *IEEE Trans. Knowledge and Data Engineering*, vol. 10, no. 6, pp. 988-995 1998.

[11] N. Dimitrova, H.-J. Zhang, B. Shahraray, I. Sezan, T. Huang, and A. Zakhor, "Applications of video-content analysis and retrieval," *IEEE Multimedia*, vol. 6, no. 2, pp. 14-17, 1999.

[12] S. Pekowsky and R. Jaeger, "The set-top box as "multi-media terminal," *IEEE Trans. Consumer Electronics*, vol. 44, no. 3, pp. 833 – 840, Aug. 1998.

[13] Y. Fu, A. Ekin, A. M. Tekalp, and R. Mehrotra, "Temporal segmentation of video objects for hierarchical object-based motion description," *IEEE Trans. Image Processing*, vol. 11, pp. 135–145, Feb. 2002.

[14] D. Zhong, and S. Chang, "Real-time view recognition and event detection for sports video," *Journal of Visual Communication & Image Representation*, vol. 15, no. 3, pp. 330-347, 2004.

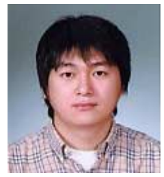
[15] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic Soccer Video Analysis and Summarization," *IEEE Trans. Image Processing*, vol. 12, no. 7, pp. 796-807, 2003.

[16] A. Ekin and A. M. Tekalp, "Generic Play-break Event Detection for Summarization and Hierarchical Sports Video Analysis," *in Proc. IEEE Int. Conf. Multimedia & Expo 2003*, pp. 27-29, 2003.

[17] M. Kumano, Y. Ariki, K. Tsukada, S. Hamaguchi, and H. Kiyose, "Automatic Extraction of PC Scenes Based on Feature Mining for a Real Time Delivery System of Baseball Highlight Scenes," *in Proc. IEEE Int. Conf. Multimedia and Expo 2004*, pp. 277- 280, 2004.

[18] S. H. Jin, T. M. Bae, Y. M. Ro, H.-R. Kim, and M. Kim, "Intelligent Broadcasting System and Services for Personalized Semantic Contents Consumption," *Expert system with applications*, submitted for publication.

Sung Ho Jin received his B.S. degree from the Dept. of Electronics and Control Engineering at Chung-Ang University, Seoul, Korea, in 2000, and his M.S. degree from the School of Engineering at Information and Communications University (ICU), Daejeon, Korea, in 2002. He is currently a Ph.D candidate at ICU. His major interests are video indexing and segmentation, image and video retrieval, multimodal, pattern recognition, TV-Anytime, and MPEG-7 based broadcasting technologies.



Jun Ho Cho received his B.S. from the Dept. of EEE at Chung-Ang University, Seoul, Korea, in 2004. Currently, he is an M.S. candidate in the Dept. of Electronics Engineering at Information and Communications University, Daejeon, Korea. His major interests are in MPEG-A, image and video retrieval, and MPEG-7 based broadcasting technologies.

Yong Man Ro received his B.S. from Yonsei University, Seoul, Korea, in 1981, and his M.S. and Ph.D. degrees from the Korea Advanced Institute in Science and Technology (KAIST), in 1987 and 1992, respectively. In 1987, he was a staff associate at Columbia University, and from 1992 to 1995, he was a visiting researcher at the University of California at Irvine and KAIST. In 1996, he was a research fellow in the Dept. of EECS at the University of California at Berkeley. In 1997 he joined Information and Communications University, Korea, where he is currently a professor and director of the Image Video System Lab. His research interests include image/video processing, MPEG-7, MPEG-21, feature recognition, image/video indexing, and spectral analysis of image signal. In 1992, he received the Young Investigator Finalist Award in ISMRM. He is a senior member of the IEEE and a member of SPIE and ISMRM.