Video Event Filtering in Consumer Domain

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

***Abstract*—A real-time content filtering system for live broadcasts in TV terminals enables a TV viewer to obtain desired video events from multiple channel broadcasts. In order to achieve stable and reliable event filtering for multiple video inputs, real-time filtering requirements such as frame sampling rate per channel, number of input channels, and buffer condition should be considered to meet the limited capacity of the terminal. In this paper, we propose a method of selecting those requirements by modeling a filtering system using a D/M/1 queue. By minimizing the queueing time of input video frames, the proposed model can maximize filtering capacity for multiple video inputs in real-time. To verify the proposed model and analysis, we perform experiments on soccer videos with the proposed model. The experimental results show that the proposed model allows for more efficient choice of requirements within limited resources.**

***Index Terms*—D/M/1 queue, real-time processing, video event filtering.**

I. INTRODUCTION

WITH the growth of digital broadcasting and the wide distribution of broadcasting devices such as set-top boxes (STBs) and personal video recorders (PVRs), the number of broadcasting channels has increased and broadcasting services have become more personalized and intelligent. As such, today's TV terminal is equipped with more advanced structures and functions such as picture-in-picture, time-shift play, channel recording, etc.

But the increasing number of channels and broadcasting contents complicate TV viewers' efforts to find their favorite broadcasts quickly and efficiently. In addition, obtaining meaningful events from live broadcasts becomes more difficult because the corresponding metadata of live broadcasts is not available [1]. In fact, most live broadcasts cannot afford to provide related metadata as it must be prepared before the broadcast is aired.

In order to provide a personalized watching environment in digital broadcasting, a meaningful event search in real-time is required on the side of the end-user, i.e., the TV terminal. The goal of real-time content filtering is to develop a user-based broadcasting service for live broadcasts that provides content-level information within the limited capacity of the TV terminal. The service scenario is as follows: a TV viewer watches his/her favorite broadcasts (e.g., a drama or sitcom) on one channel, while the other channels broadcast other programs (e.g., the final round of a soccer game) that contain interesting events (e.g., shooting or goal events). To locate events of interest on other channels, a real-time filtering technique, which recognizes and extracts meaningful content of the live broadcast, should be embedded in the TV terminal.

However, current broadcasting systems can provide a content-level indexing service with pre-made metadata, but do not offer meaningful event searching with respect to live broadcasts except simple program guiding services with electronic program guides [2], [3]. In [4], [5], personalized broadcasting systems have been implemented for user terminals but they also employ pre-made metadata which contains video segmentation and indexing information.

So far various state-of-the-art techniques for analyzing videos have been applied to video archiving systems for video summarization, video segmentation, content management, and metadata authoring of broadcasts [6]–[11]. These techniques are only performed on the side of the service providers or the content creators. In [6], C.-W. Ngo *et al.* proposed a unified approach for video summarization by using graph-based scene modeling and highlight detection. The resulting clusters formed a directed temporal graph and a shortest path algorithm was proposed to efficiently detect video scenes. X. Zhu *et al.* [7] presented a hierarchical video content description and summarization strategy by adopting video content description ontology. Y. Li *et al.* [8] addressed a content-based movie parsing and indexing approach through the analysis of both audio and visual sources and accounting for their interrelations to extract high-level semantic cues. In the references of [9], [10], the research approaches focused on real-time processing for event detection. By proposing a defined dynamic threshold model enabling the capture of the statistical properties of the scene changes [9] or developing a new pipelined digital video processing architecture for digitizing, processing, indexing, and compressing video [10], they showed the significant improvement of scene-detection accuracy. In [11], various kinds of video analysis techniques such as event detection, video segmentation and shot detection were introduced and summarized. Most of the previous approaches showed good performances to detect segmentation information of video sequences but were limitedly developed for the service providers or the content creators.

In the end-user's terminal, several approaches have been performed [12]–[17]. In the approaches shown in [12]–[14], the semantic information of videos such as shots and scenes were extracted and detected after video recoding. A few works [15]–[17] had performed to extract and filter video segments in real-time. However they were limited to detect TV commercial breaks which were inserted within a main broadcast. N. Dimitrova *et al.* [15] researched video analysis algorithms

Manuscript received June 9, 2006; revised June 15, 2006. This work was supported in part by Ministry of Information and Communication, Republic of Korea, under SmarTV project.

The authors are with the Image and Video System Lab, Information and Communications University (ICU), Daejeon 305–732, Republic of Korea (e-mail: wh966@icu.ac.kr; yro@icu.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2007.909203

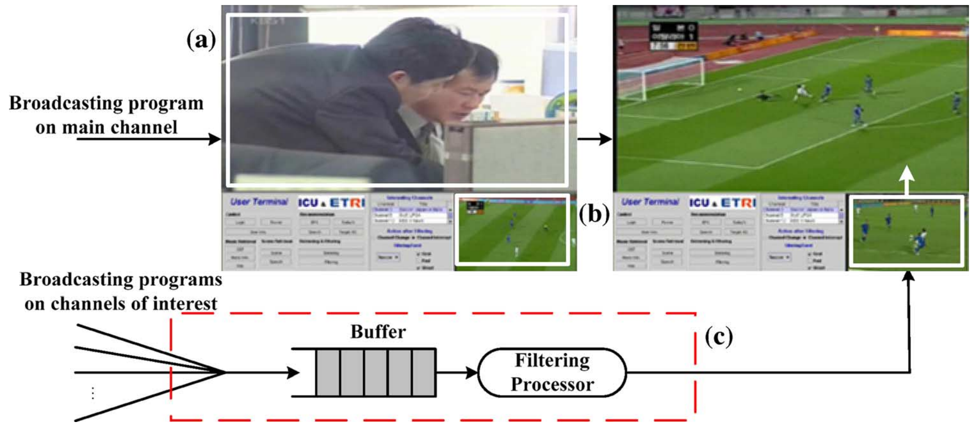


Fig. 1. Conceptual consumer system for real-time video event filtering: (a) main display for a program to be watched, (b) simulated display for the second program to be filtered, and (c) filtering processor.

and architecture for abstracted video representation in consumer domain applications and also they developed a tool for commercial skipping through the detection of black frames and changes in activity [16], [17]. Our work, however, is focused on establishing a system that enables the indexing and analyzing of live broadcasts at the content level.

In [18], we proposed real-time content filtering for multiple live broadcasts. In addition, system requirements for stable filtering were developed by using a D/D/1 model, which has deterministic inter-arrival and service patterns. Even though the previous model guarantees stable filtering by assuming the worst case of the buffer where frames with the longest service time are successive, there is room for selecting requirements more optimally. With the aid of the available filtering system, this paper will attempt more practically to model the proposed system for the provision of more appropriate filtering requirements. This paper focuses on finding and modeling the filtering system with D/M/1 queue model which could increase available system requirements such as the number of input channels and the frame sampling rate. Furthermore, we compare the system requirements between D/D/1 and D/M/1 models.

The rest of this paper is organized as follows. We first give an overview of the proposed system and general filtering procedure in Section II, and Section III offers a filtering system analysis based on the D/M/1 model. Section IV presents experiments in which soccer videos are applied to the proposed system and verifies the proposed model with the experimental results of five soccer videos. Concluding remarks are drawn in Section V.

II. Video Event Filtering in Consumer Terminals

Before modeling the proposed filtering system, we will present an overview of the system and its filtering algorithm.

A. *Overview of Broadcasting Content Filtering System*

For the proposed real-time content filtering, it is assumed that TV terminals equipped with a STB or PVR will evolve into multimedia centers in the home with computing capability and home server connections [19], [20]. Fig. 1 shows a conceptual TV terminal with a real-time content filtering function which receives multi-input. The terminal requires a digital tuner enabling it to extract each broadcasting stream time-division, or multiple tuners for the filtering of multiple channels. To meet the case in which it receives multiple broadcasts from the channels selected for filtering as well as the main channel, the terminal of Fig. 1 should have more than two TV tuners: one tuner accepts a broadcast stream from the main channel to be watched and the other receive streams from the selected channels to be filtered. This work also supposes that the receiver components, i.e. the demux and decoder, before the buffer for the filtering processor, are able to support the decoding of multi-streams.

For a TV user to acquire video events of interest from selected channels while at the same time watching a broadcast of the main channel, the system should have a filtering processor which can detect the events by content-level analysis. According to many previous works [21], it is difficult to show high performance of detecting video events in a compressed domain because the spatial and temporal features of the domain are insufficient to identify compressed video frames. Therefore, we perform event filtering with decoded video frames.

In the filtering processor, meaningful video events are extracted from input broadcasts and filtered results are sent to a display unit. As shown in Fig. 1, the filtering processor monitors channels and filters meaningful video events according to the TV viewer's preference or intention. As soon as the terminal finds the selected events—that is, the viewer's desired events filtered from his/her channels referred to as "video events of interest"—it gives a notification of this fact by means of the channel change or channel interrupt function. In current TVs, the picture-in-picture function can be utilized to effectively show the filtering results.

The system needs a buffer for queueing multiple video frames fetched from channels. Thus, certain filtering criteria that simultaneously prevent buffer overflow and keep filtering performance are applied. For example, the number of allowable input channels required to guarantee real-time filtering depends on the frame sampling rate, buffer condition, and etc. (these are discussed in detail in Section III). Finding these criteria is the goal of this work.

B. *Generic Filtering Processor With Visual Features*

The general structure of sports videos is that they possess repetitive occurrences of meaningful events and are very fre-

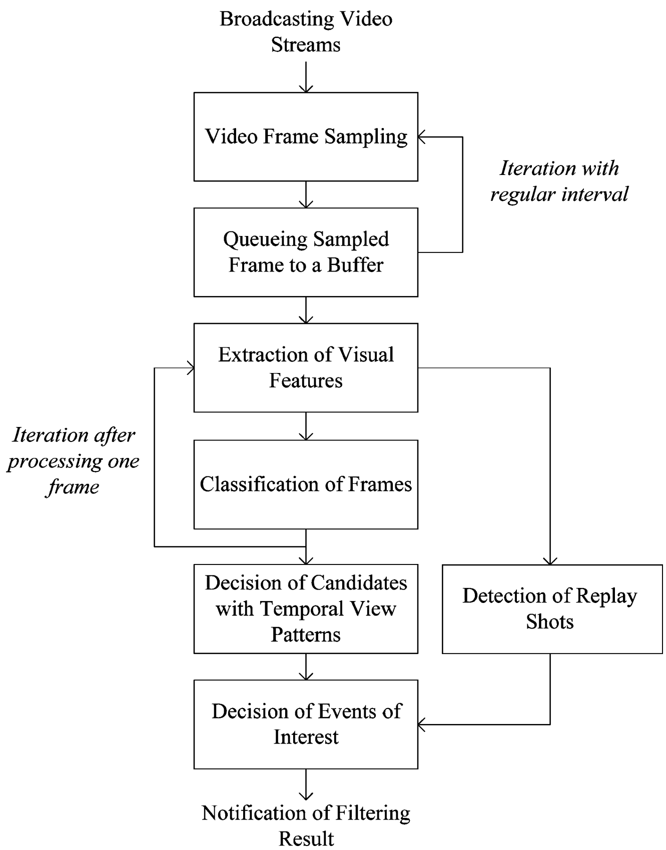


Fig. 2. Generic filtering processor for proposed filtering.

quently broadcast on the spot, making the proposed real-time filtering in a TV terminal suitable for such videos. In order to represent such broadcasts, there are two approaches. The first uses object-based features (e.g., object motions) in sports video indexing to enable high-level domain analysis. However, its disadvantage is that extraction may be computationally expensive for real-time implementation [22], [23]. To minimize computational complexity we employ a second approach that uses cinema features, made up of common video composition and production rules (e.g., replay), as well as the threshold methods used in previous works [24]–[26].

In this work, we use a generic filtering algorithm for sports videos based on cinema features. To extract video events, spatial features such as edges and colors, and the temporal patterns of frames, are used for the generic filtering processor. Fig. 2 shows a flowchart of the generic filtering processor used for the proposed filtering system.

By enabling the filtering function by user intention or an agent of the TV terminal, channels of interest are determined. A filtering processor then samples input videos from the selected channels at regular intervals and queues the obtained frames in a buffer. In the order of FCFS (first-come, first- served), visual features such as color, edges and motions are extracted from them. Using the given features, it classifies the types of frames which are defined depending on the kind of input video. The processor decides the candidates of the events of interest by acquiring temporal patterns with the classified frames. Our work is based on the fact that a meaningful video event tends to have a specific temporal flow by successive frames. According to our observations, meaningful events in sports videos show a sequence consisting of several different camera views. In most cases, the views of main cameras show the outline of a sport with the recurrence of scoreboards or captions, whereas the views of other cameras show short replays without those items. To verify the results of the event filtering, the proposed filtering uses replay shot detection and then decides the notification for the filtering result.

In the above flowchart, the time of iteration between the feature extraction and frame classification depends on the type of input frame because different frame types have different features. For example, in a soccer video, the dominant color feature is important to detect the ground view, but is less important in a crowd or person view. Thus, the proposed filtering processor shows different frame classification times, thus avoiding the overflow problem usually caused at the buffer.

III. Modeling With D/M/1 Queue

As shown in Figs. 1 and 2, the proposed filtering system accepts video frames sampled regularly from multiple channels. The view type of the frame obtained from one channel is independent of those of frames from other channels. The fact reveals that each filtering job is statistically independent in the filtering process. Because the filtering jobs are performed with the frames buffered according to FCFS discipline, it is impossible that more than two filtering jobs occur simultaneously. So the filtering process can be treated as a counting process. Furthermore one can observe the fact that the proposed filtering processor (shown in Fig. 2) has different filtering service times depending on the view types of the sampled frames (shown in Table II).

Given that the view types of the frames in a buffer are statistically independent, follow a counting process, and have different service times (filtering times), the number of filtering occurrences within a certain time interval becomes a Poisson random variable. Then, the probability distribution of the service time is an exponential distribution. We apply a D/M/1 queue model referred from [27], [28] to analyze the proposed filtering system, which covers constant inter-arrivals to a single-server buffer with exponential service time distribution, unlimited capacity and FCFS queueing discipline.

As shown in Fig. 3, let a video frame arrive at time $t_0 = 0, t_1, t_2, \ldots$ and get served by a single filtering processor. Let $T_n = t_{n+1} - t_n, n = 1, 2, 3, \ldots$, be independent and identically distributed random variables with a distribution $A(\cdot)$ with mean $1/\lambda$. From the characteristics that the filtering occurrences can be expressed as a form of Poisson random variable, we let the service time S_n follows an exponential distribution with mean $1/\mu$. Here, λ and μ denote the inter-arrival rate and the service rate of an input frame, respectively.

First, we define the traffic intensity as

$$\rho = \frac{\lambda}{\mu}. \tag{1}$$

To maintain the stability of the process, the traffic intensity ρ should be kept at less than one, where it is the first criterion for

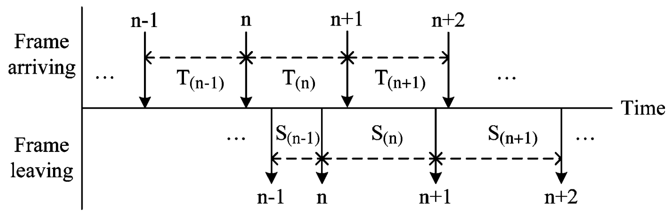


Fig. 3. Successive frames in a D/M/1 queue model.

stability. In the proposed filtering process with multiple inputs, the inter-arrival rate λ is given by

$$\lambda = N \cdot f \tag{2}$$

where N indicates the number of input channels and f is the frame sampling rate per channel.

Service rate μ takes the form of

$$\mu = \frac{1}{\frac{1}{C_{total}} \sum_{i=1}^{M} C(v_i) \cdot PT(v_i)} \tag{3}$$

where v_i denotes a frame view type defined in input video and M is the total number of view types. $C(v_i)$ and $PT(v_i)$ represent the number of frames with view type v_i and its processing time, respectively. C_{total} is the total number of input frames.

Then, (1) and (2) yield

$$N \cdot f \cdot \frac{1}{C_{total}} \sum_{i=1}^{M} C(v_i) \cdot PT(v_i) < 1. \tag{4}$$

The second criterion can be established from the trade-off between the frame sampling rate and filtering performance. The filtering performance depends on the frame sampling rate and tends to decrease as it decreases. Thus, the sampling rate should be selected appropriately to satisfy the minimum filtering performance that TV users can accept. The minimum performance of the filtering system should be maintained during changes in the frame sampling rate or the number of channels. Let $FP(f)$ be a function which represents the relationship between the filtering performance and the frame sampling rate. Then, a frame sampling rate f_s with which a filtering system meets the minimum performance is as follows:

$$\begin{aligned} &f_s = \arg\min_f \{FP(f) \geq T_{fp}\}, \\ &f > f_s \end{aligned} \tag{5}$$

where T_{fp} is a threshold value of the filtering performance.

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

For the D/M/1 model, however, a video frame with long service time can make subsequent video frames wait for service. This enables the model to have a buffer for those frames, and the expected buffer length is estimated through experiments.

In Fig. 3, X_n is the number of customers (i.e., input video frames) in the system just before the nth arrival, and B_n is the number of customers served during inter-arrival period T_n between the nth and the $(\mathrm{n}+1)$st arrival. Given that the arrival process is deterministic, it is given by

$$a(t) = \begin{cases} 1 & if\ t = \frac{1}{\lambda} \\ 0 & otherwise \end{cases} \tag{6}$$

where t is the time for the arrival of one customer.

Since $A(t)$ is the cumulative distribution function (CDF) of $a(t)$ and B_n follows Poisson distribution, we introduce the simplifying notation:

$$\begin{aligned} b_j &= P(j\ services\ during\ T) \\ &= P(B=j) \quad j = 0,1,2,\ldots \\ &= \int_0^\infty P(B=j|T=t)dA(t) \\ &= \int_0^\infty e^{-ut}\frac{(ut)^j}{j!}dA(t) \quad (t \geq 0). \end{aligned} \tag{7}$$

If the inter-arrival times are assumed independent, and the service time is assumed exponential, the time-dependent script is dropped.

Letting $\beta(z)$ be the probability generating function of b_j, then $\beta(z)$ is simplified as

$$\begin{aligned} \beta(z) &= \sum_{n=0}^{\infty} b_n z^n \\ &= \int_0^\infty e^{-\mu t}\sum_{n=0}^{\infty}\frac{(\mu t z)^n}{n!}dA(t) \\ &= \int_0^\infty e^{-\mu t(1-z)}dA(t) = A^*\left[\mu(1-z)\right]. \end{aligned} \tag{8}$$

where $A^*(z)$ is the Laplace-Stieltjes transform of $A(t)$.

Since it is shown that $z = \beta(z)$ in [27], (8) can be written as

$$z = A^*\left[\mu(1-z)\right] \tag{9}$$

By using (6), we get

$$\begin{aligned} A^*(z) &= \int_0^\infty e^{-zt}dA(t) \\ &= \int_0^\infty e^{-zt}a(t)dt \\ &= e^{-z/\lambda} \quad if\ t = 1/\lambda. \end{aligned} \tag{10}$$

Thus,

$$\begin{aligned} z &= A^*\left[\mu(1-z)\right] \\ &= e^{-(\mu-\mu z)/\lambda} \\ &= e^{-(1-z)/\rho}. \end{aligned} \tag{11}$$

Note that σ is the root of $z = A^*[\mu(1-z)]$. In most cases, the root σ is determined by numerical techniques.

From [27], the length of the expected buffer is given by

$$L_q = \frac{\rho\sigma}{1-\sigma}. \tag{12}$$

To apply a buffer into the filtering system, we should use a round-up of the expected buffer length,

$$M_q = \langle L_q \rangle. \tag{13}$$

IV. Experimental Results

In this section, experiments are performed with soccer videos to demonstrate the proposed queue model and to find the requirements of content filtering. Five soccer videos are used for the test, each of which is about 90 minutes in length. In order to evaluate the filtering requirements within the given resources, we used three different terminals: P_1 with 650 MHz CPU and 128 MByte memory; P_2 with 1.7 GHz and 512 MByte; and P_3 with 3 GHz and 2 GByte, respectively. Actually, the terminal with 650 MHz CPU shows computing power which is similar to currently available STBs.

A. Filtering on Soccer Videos

Shooting events including goals are utilized as video events of interest in soccer videos. Before extracting the events, frame view types in soccer are divided into four categories: 1) global view with goal post (v_1); 2) global view without goal post (v_2); 3) medium view (v_3); and 4) close-up view (v_4). As described in Fig. 2, the filtering algorithm used in this work to detect the shooting events consists of frame view classification and replay shot detection. A temporal pattern by frame view sequence decides the candidates of shooting events and then the occurrence of the replay shot verifies the detection of a shooting event. Fig. 4 shows the temporal patterns observed as features of the events in this work: c_1 is the sequence of global views with goal post, c_2 is the view transition, c_3 is the sequence of either close-up views or medium views, c_4 is the transition describing the occurrence of the replay, and c_5 represents the sequence of global views with or without goal post.

It is observed that the view transition from the global view with goal post to the close-up view or medium view happens at the beginning of a meaningful event in sports broadcasts. As well, a replay shot follows just after a meaningful event.

For classifying frame view types, we extract one dominant color of the HSV domain from the sub-blocks of a video frame, which represents the color of the playing field. A Sobel operator with a 3 × 3 mask is used to detect edges and a non-maxima suppression method is applied to thin them out. In this work, Hough transform makes it easy to find the layout of the playing field which can be represented by lines of a polar coordinate. In order to determine the events of interest from the candidates consisting of frame sequences, we analyze the continuity of the scoreboard within the sequences and check the appearances of replay shots. As text characters in a frame form a regular texture containing vertical and horizontal strokes [29], input frames are compared with the reference frame which represents the scoreboard. (A more detailed algorithm for soccer video is described in our previous paper [18].)

We established a TV terminal playing a video on one channel, while another video from another channel was being filtered using the proposed filtering system. In the experiments, we measured the consuming time and filtering performance, and obtained the filtering requirements by applying them to the proposed queue model. Table I describes the number of input frames which are counted from the test videos according to frame views.

In Table II, the processing time of each frame view, within which a single input frame is served in the proposed filtering system, is shown for the terminals.

In addition, the performance of the proposed view decision, with three frames per second sampling rate and one channel of interest, is observed and the following results derived. The

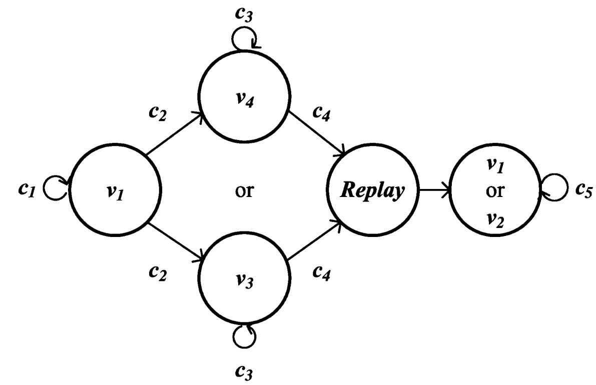


Fig. 4. Temporal view patterns for shooting events in soccer videos.

TABLE I
Number of Frames Depending on Frame Views

	$C(v_1)$	$C(v_2)$	$C(v_3)$	$C(v_4)$	C_{total}
Number of frames	53680	79920	30285	51380	215265

TABLE II
Processing Times of Frame Views Measured in Each Terminal

Time \ Terminal	P_1	P_2	P_3
$PT(v_1)$	0.340 sec.	0.206 sec.	0.110 sec.
$PT(v_2)$	0.281 sec.	0.174 sec.	0.088 sec.
$PT(v_3)$	0.270 sec.	0.075 sec.	0.045 sec.
$PT(v_4)$	0.170 sec.	0.037 sec.	0.025 sec.

total average recall rate was over 93.5% and the total average precision rate was approximately 89.5%. The filtering performance of the shooting events (including goal events) showed an average recall rate of 81.5% and an average precision rate of 76.4% was shown in the test videos. The detection of goal events is more accurate than that of shooting because the goals have replay shots without exception.

B. Filtering Requirement Analysis Using Proposed D/M/1 Model

Substituting the experimental results shown in Tables I and II to (4), the first criterion of the soccer video, i.e. its stability in real-time for the proposed filtering system, can be found as,

$$N \cdot f < 1/\frac{1}{C_{total}} \sum_{i=1}^{M} C(v_i) \cdot PT(v_i). \tag{14}$$

Since the right term of inequality is calculated by using the results of Tables I and II, available N and f in each terminal are obtained as shown in Fig. 5.

Comparing the proposed D/M/1 model with our previous model (D/D/1) in each terminal, Fig. 5 shows that the proposed model gives more wide-ranging choices with regard to the number of channels and frame sampling rate. With 2.5 frames per second, the D/D/1 model gives 1.17, 1.94, and 3.64 for P_1, P_2, and P_3 as the maximum number of available channels, respectively. Otherwise the D/M/1 model shows 1.49, 2.96, and 9.52. As expected, the results prove that an exponential service model is better than a deterministic one in terms of finding suitable filtering factors.

The second criterion of (5) is then verified by evaluating the filtering performance of the proposed filtering algorithm. Fig. 6 presents the change of filtering performance on the test soccer videos by frame sampling rate whose values determine the minimum frame sampling rate. As shown, the performance (recall rate in the figure) decreases as the frame sampling rate decreases.

From Fig. 6, it is shown that the maximum permissible limit of sampling rate is determined by the tolerance T_{fp} of the filtering performance. In the case that the system permits 80% filtering performance of T_{fp}, it is observed that the frame sampling rate f_s becomes 2.5 frames per second by the experimental result.

With the two criteria obtained above, we acquire the filtering requirements for the filtering system used on soccer videos. As shown, the number of input channels depends on both frame sampling rate and terminal capability. By assuming the confidence limit of the filtering performance T_{fp} as 80%, we also get the minimum sampling rate from Fig. 6.

In Fig. 7, line L_1 and L_2 indicate the conditions of (14) and Fig. 6, respectively. Here, line L_1 is the number of available channels for D/M/1 model shown in Fig. 5. Line L_2 denotes the value of the frame sampling rate obtained from Fig. 6. According to the sum of processing times with regard to input frames, the lower region of line L_1 shows that the number of input channels is inversely proportional to the frame sampling rate. The right region of line L_2 is the range of sampling rate required to maintain over 80% filtering performance. And line L_3 represents the minimum number of input channels, i.e., one channel.

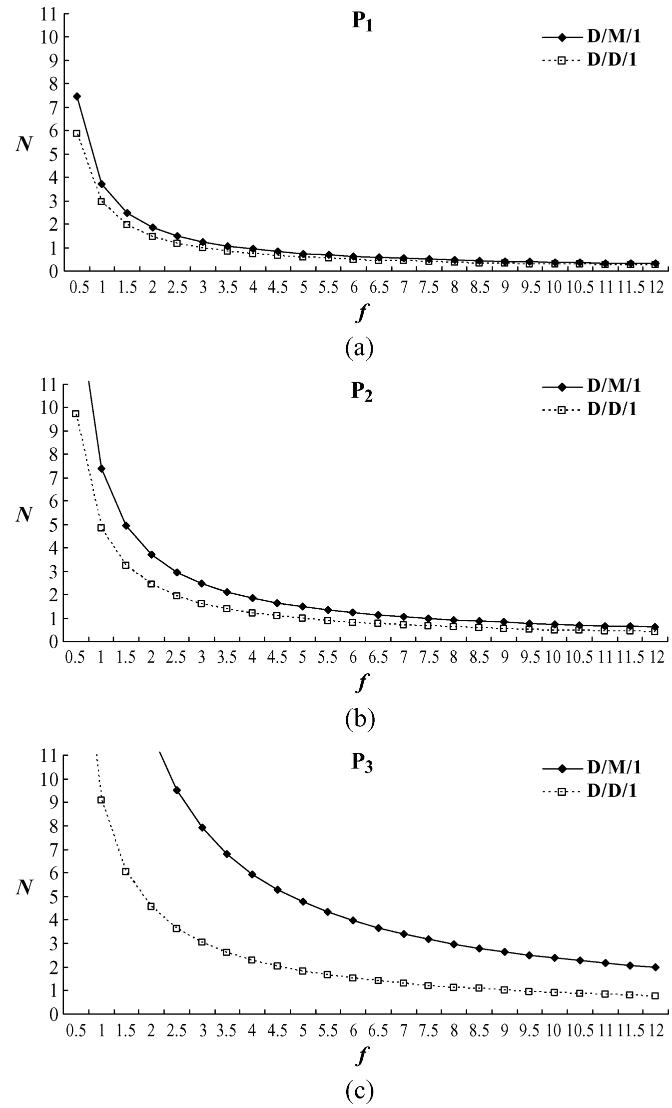


Fig. 5. Number of available channels according to variation of frame sampling rate in (a) P_1, (b) P_2, and (c) P_3.

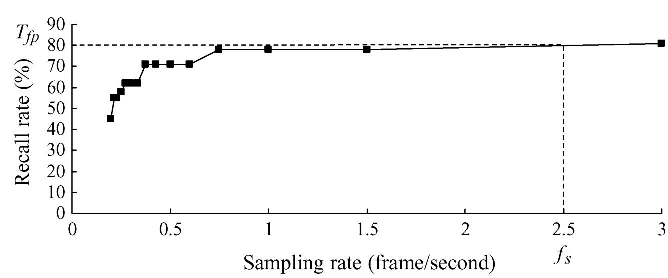


Fig. 6. Variation of filtering performance according to frame sampling rate.

To keep stability in the filtering system, therefore, the number of input channels and the frame sampling rate should be selected in the range where the three criteria by line L_1, L_2, and L_3 meet. Supposing that the confidence limit of the filtering performance is 80%, Fig. 7 illustrates the following results: one input channel is allowable for real-time filtering in Terminal 1 at sampling

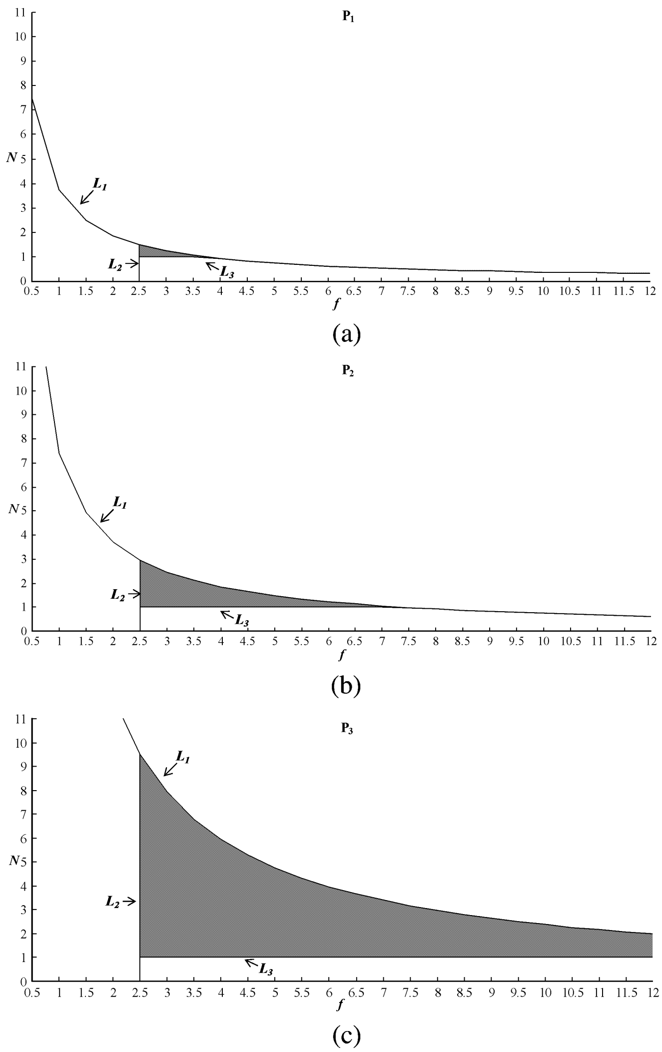


Fig. 7. Regions including requirements that meet the two criteria of the proposed D/M/1 model on soccer videos in (a) P_1, (b) P_2, and (c) P_3.

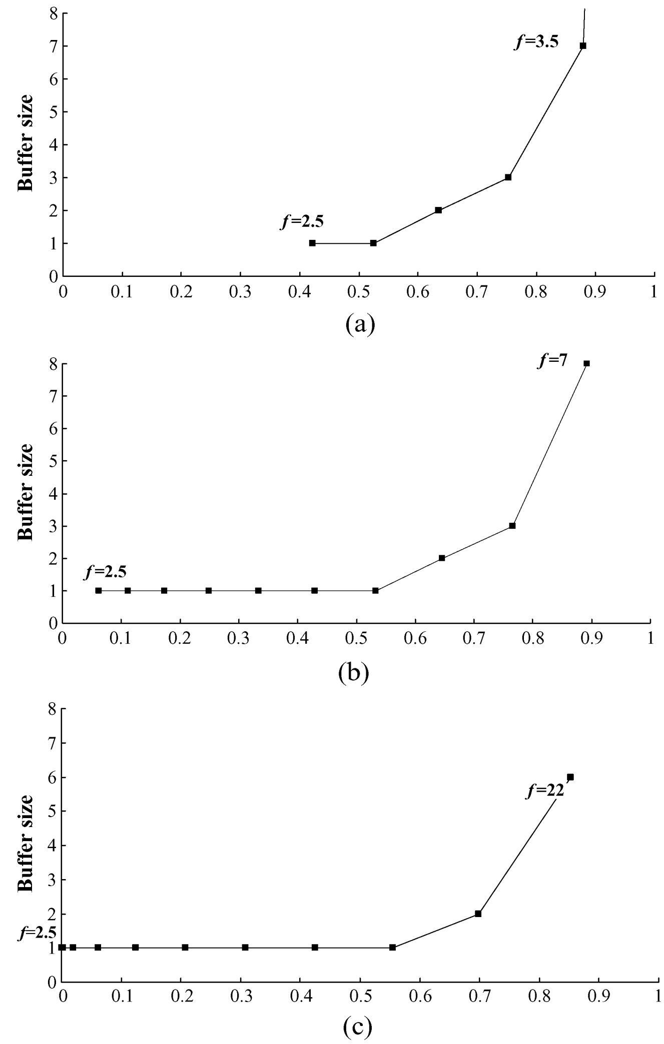


Fig. 8. Buffer lengths expected and used in the proposed D/M/1 model in (a) P_1, (b) P_2, and (c) P_3.

rates between 2.5 and about 3.8 frames per second. In Terminal 2, one or three channels are allowable at sampling rates between 2.5 and 7.5 frames per second. Terminal 3 can use nine channels maximum.

In contrast to the D/D/1 model, the proposed D/M/1 model needs a buffer for input frames as a trade-off of extended requirement selection. In Fig. 8, buffer size M_q is drawn for each terminal. Through the results, it is verified that P1 needs more buffer size compared to the others; in other words, the terminal with higher system capacity performs the proposed filtering process more quickly. In order to realize the proposed filtering system, these experimental results should be referred with the stability criteria.

So far, we have applied the proposed D/M/1 model to soccer videos so that we can verify that it offers a more efficient and practical approach for a real-time content filtering system, even though it needs a buffer as a trade-off.

V. Conclusions

In this paper, we proposed a real-time content filtering system model for live broadcasts, which provides personalized video events in TV terminals, to select suitable filtering requirements. Compared with the previous model with D/D/1 queue, the effectiveness of the proposed model has been verified. By applying a D/M/1 queue, it is shown that the proposed system is suitable for real-time filtering of live broadcasts and permits wider choices of filtering requirements. Our experimental results revealed that even if the D/M/1 model needs a buffer for queueing input frames as a tradeoff, it is more efficient in terms of choices than a D/D/1 model. In addition, the results show that even a low-performance terminal with 650 MHz CPU can perform the filtering function in real-time. Thus, the proposed system model and its requirements can confirm that the real-time filtering of live broadcasts is possible with currently available set-top boxes. Thus, we conclude that the proposed filtering system can be established to provide a content-level broadcasting service in the consumer domain.

For future work, the filtering algorithm requires more enhanced filtering performance with real-time processing. In addition, it is necessary that the algorithm can be extendable to other sport videos such as baseball, basketball, golf, etc., and that it can be applied in a real environment.

References

[1] TVAF, "Phase 2 Benchmark Features," SP001v20, 2005, p. 9 [Online]. Available: http://www.tv-anytime.org

[2] 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.

[3] 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.

[4] K. Kang, J.-G. Kim, H. K. Lee, H. S. Chang, S.-J. Yang, Y.-T. Kim, H. K. Lee, and J. Kim, "Metadata broadcasting for personalized service: A practical solution," *ETRI Journal*, vol. 26, no. 5, pp. 452–466, 2004.

[5] H. K. Lee, J.-G. Kim, S.-J. Yang, and J. Hong, "Personalized TV services based on TV-anytime for personal digital recorder," *IEEE Trans. Consumer Electronics*, vol. 51, no. 3, pp. 885–892, 2005.

[6] 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.

[7] 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.

[8] Y. Li, S. Narayanan, and C.-C. Jay Kuo, "Content-based movie analysis and indexing based on audiovisual cues," *IEEE Trans. Circuits and System for Video Technology*, vol. 14, no. 8, pp. 1073–1085, 2004.

[9] 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.

[10] J. M. Gauch, S. Gauch, S. Bouix, and X. Zhu, "Real time video scene detection and classification," *Information Processing and Management*, vol. 35, pp. 381–400, 1999.

[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. 9, no. 3, pp. 42–55, 2002.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] N. Dimitrova, H. Elenbass, T. McGee, and L. Agnihotri, "An architecture for video content filtering in consumer domain," in *Proc. Int. Conf. on Information Technology: Coding and Computing 2000*, 2000, pp. 214–221.

[17] 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.

[18] S. H. Jin, J. H. Cho, and Y. M. Ro, "Meaningful scene filtering for TV terminals," *IEEE Trans. Consumer Electronics*, vol. 52, no. 1, pp. 263–267, 2006.

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

[20] J.-C. Moon, H.-S. Lim, and S.-J. Kang, "Real-time event kernel architecture for Home-Network Gateway Set-Top-Box (HNGS)," *IEEE Trans. Consumer Electronics*, vol. 45, no. 3, pp. 488–495, 1999.

[21] Y. Li and C.-C. Jay Kuo, *The Video Content Analysis Using Multimodal Information: For Movie Content Extraction, Indexing and Representation*. Massachusetts: Kluwer Academic Publishers, 2003, ch. 3.

[22] 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, 2002.

[23] 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.

[24] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Trans. Image Processing*, vol. 12, no. 7, 2003.

[25] 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*, 2003, pp. 27–29.

[26] 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*, 2004, pp. 277–280.

[27] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*. New York: John Wiley & Sons, 1998, ch. 1, 2, 6.

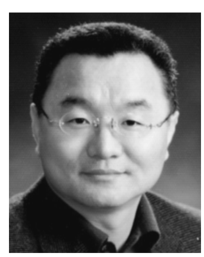
[28] L. Kleinrock, *Queueing System*. New York: Wiley, 1975, ch. 1, 2.

[29] C. Wolf, J.-M. Jolion, and F. Chassaing, "Text localization, enhancement and binarization in multimedia documents," in *Proc. 16th Int. Conf. Pattern Recognition*, 2002, vol. 2, pp. 1037–1040.

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 and Ph.D. from the School of Engineering at Information and Communications University (ICU), Daejeon, Korea, in 2002 and 2006, respectively.

He worked as an intern researcher for Spintech Co., Ltd in 2001. In 2003 to 2004, he joined Inter-Information Co. Ltd as well as Electronic Telecommunications Research Institute, Korea as an invited researcher. Currently, he is a Post Doc. Researcher in ICU. His major interests are video indexing and segmentation, image and video retrieval, multimodal, pattern recognition, MPEG-A, TV-Anytime, and MPEG-7 based broadcasting technologies.

Dr. Sung Ho Jin is a member of IEEE.



Yong Man Ro received the B.S. degree from the Department of Electronics Engineering, Yonsei University, Seoul, Korea, in 1985 and the M.S. and Ph.D. degrees from the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), in 1987 and 1992, respectively.

He was a Researcher at Columbia University, New York, in 1987 to 1988, a Visiting Researcher at the University of California at Irvine and KAIST from 1992 to 1995, and a Research Fellow in the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in 1996. He was also a Professor in the Department of Computer Engineering, Daejeon University, until 1996. Since 1997, he has been with the Information and Communications University (ICU), Korea, where he is currently a Professor of the Multimedia Communication Group and the director of IVY (Image/Video System) Lab. He participated in international standardizations including MPEG-7, MPEG-21, MPEG-A, and SVC, where he contributed several MPEG-7, MPEG-21, MPEG-A, and SVC standardization works. His current research interests include multimedia adaptation/modality conversion, multimedia information retrieval, medical image processing and watermarking.

Prof. Yong Man Ro is a senior member of IEEE as well as member of ISMRM and SPIE. He received the Young Investigator Finalist Award of ISMRM in 1992.