An Experimental Analysis of Accumulated Audience's Comments for Video Summarization

Yoshia Saito [*], Yoshiki Isogai [*] and Yuko Murayama [*]

[*] Graduate School of Software and Information Science, Iwate Prefectural University, Japan
{y-saito, murayama}@iwate-pu.ac.jp, y.isogai@comm.soft.iwate-pu.ac.jp

Abstract - In this paper, we propose an audience-oriented video summarization scheme on video sharing services. The proposed scheme analyzes audiences' feedbacks such as rating and comments in a video and finds important scenes where there are a lot of feedbacks from the audiences. Then, the video is summarized by collecting the important scenes from audiences' point of view although typically it is summarized from video producers'/providers' point of view. As the first step toward the audience-oriented video summarization, we focus on comments as the audiences' feedbacks because currently some video sharing services allow audiences to comment on a specific scene storing their playback time. We assume there is a relationship between the number of audiences' comments on a scene and importance of the scene because the comments represent audiences' willingness to watch the scene. We report an experimental analysis for verification of the hypothesis and discuss some solutions to realize audience-oriented video summarization taking into account the experiment results.

Keywords: Internet broadcast, video sharing service, audiences' feedback, comments, audience-oriented video summarization.

1 INTRODUCTION

In recent years, most Internet users have broadband Internet connections and multimedia contents become popular on the Web. There are a lot of video sharing services nowadays such as YouTube [1] and Yahoo! Video [2]. A huge number of videos are shared and hundreds of thousands of new videos are uploaded every day. It is, however, difficult for audiences to find interesting videos quickly even if they retrieved dozens of candidates by appropriate keywords since it is required to watch the videos taking long time. A solution to the issue is to provide summarized videos.

Automatic generation of video summarization techniques have been studied by a lot of researchers [3-5]. In these studies, summarization is typically realized by understanding object and event in the video and selecting important scenes. Since these studies do not get directly feedbacks from audiences and there are a lot of audiences who have different feelings, it is difficult to keep interest factors of original video for the audience. To provide attractive summarized videos for the audiences, the video summarization should be audience-oriented. That means audiences' feedbacks should be applied to the video summarization algorithm to find scenes where the audiences get interested.

Meanwhile, most video sharing services have functions to receive feedback from audiences such as rating and comments. The received feedbacks are stored in a database and available for analysis of the videos. It would be possible to find scenes where the audiences pay attention by utilizing the feedbacks. Some video sharing services allow audiences to comment on a specific scene storing their playback time. Since each feedback is related with a specific scene, the feedbacks can be used as metadata about the scenes. Thus, current video sharing services already have good database to realize audience-oriented video summarization.

In this paper, we propose an audience-oriented video summarization scheme on video sharing services. The proposed scheme analyzes audiences' feedbacks in the video and finds scenes where there are a lot of feedbacks from the audiences. Then, the video is summarized by collecting the important scenes from audiences' point of view. As the first step toward the audience-oriented video summarization, we focus on audiences' comments as the audiences' feedbacks. We assume there is relationship between the number of audiences' comments on a scene and importance of the scene for video summarization because the comments represent audiences' willingness to watch the scene. To verify the assumption, we conduct an experiment collecting ten thousand comments per video from a video sharing service and discuss whether it is possible to make a summarized video utilizing the audiences' comments.

The remainder of this paper is organized as follows. In Section 2, we describe related work. Section 3 illustrates a model of audience-oriented video summarization on a video sharing service and describes a hypothesis. In Section 4, we conduct experiments for preliminary analysis and show the results. In Section 5, we discuss solutions to realize audience-oriented video summarization taking into account the experiment results. Section 6 gives some conclusions with a brief summary and future work.

2 RELATED WORK

We can save our time by summarized video and highlight video. Recently, we can also give feedback to watched videos and share our experience. In this section, we explain difference between the summarized video and highlight video and also describe scene extraction techniques which use audiences' feedbacks.

ISSN1883-4566

2.1 Summarization and Highlight

There is difference between summarization and highlight. We define the summarized video and highlight video by reference to typical researches [6-9] as follows:

- **Summarized video** shows the story of a video content in short time.
- **Highlight video** shows a set of interesting scenes of a video content in short time.

The motivation of our research is to provide short videos so that audiences can find objective video and grasp course of story of the videos quickly. We focus on the video summarization.

2.2 Audiences' Feedbacks

There are several scene extraction techniques which use audiences' feedbacks. In [10], audiences' browsing log such as "PLAY", "STOP", "PAUSE" and "JUMP" are used for the video summarization. The audiences unintentionally give their understanding of the video to the system through the browsing operations. They measure the subjective interestingness and importance using the browsing log. In sport videos, there is a technique [11] to use audiences' reactions such as cheering and applause. The proposed technique recognize audio signal in the sport videos and extracts interesting events for the video summarization.

A concept of time-tagging is proposed in [12]. Audiences can add time-tags to videos and these tags can be used as bookmarks. It is also applied to video summarization technique by analyzing the shared time-tags and scoring the tagged segments. Current video sharing and live streaming services provide feedback functions for audiences. Several video sharing services such as YouTube and Yahoo! Video has comment and rating functions. Audiences can submit text messages to the videos and rate the videos by 5-point scale. Most live video streaming services such as Ustream.tv [13] and Stickam [14] have a chat function. In these services, audiences can send chat messages among the audiences and its broadcaster in real-time. Nico Nico Douga [15] is a video sharing service in Japan and allows audiences to comment on a specific scene storing their playback time. The comments are displayed on the video field synchronized with the commented scene as if chatted with other audiences in real-time. Since the comments correspond with specific scenes and can be easily gotten them, we use the comment data in the Nico Nico Douga for our research.

3 AUDIENCE-ORIENTED VIDEO SUMMARIZATION SCHEME

The purpose of the audience-oriented video summarization is to provide summarized videos which keep interest factors of the original ones to audiences. In this paper, the "*audience-oriented*" means utilizing feedbacks from audiences as much as possible to provide a service from audiences' view of point. The audience-oriented service would improve audience's satisfaction since it directly reflects the feedbacks.

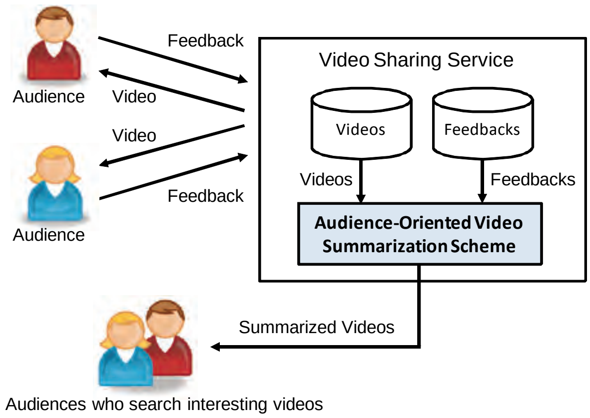

Figure 1: A model of video sharing service with audience-oriented video summarization scheme.

3.1 Overview

Figure 1 shows a model of video sharing service with the audience-oriented video summarization scheme. In this service model, a service provider delivers videos to audiences and the audiences can give feedbacks to the service provider. The feedbacks are stored in a database of the service provider. When there are audiences who search interesting videos and have several candidates to watch, the service provider generate summarized videos of the candidates applying the feedbacks appropriately. The service provider offers the summarized videos to the audiences. The audiences can decide to watch a video by reference to the summarized video. If there are not enough audiences' feedbacks for the summarization, the videos are summarized by audio-visual video summarization techniques cooperatively.

3.2 Methodology

In this paper, we use audiences' comments which are associated to specific scene as the feedbacks. In order to study an algorithm for the audience-oriented video summarization, we have a simple hypothesis about relationship between video summarization and the audiences' comments. The hypothesis is as follows:

> *There is a relationship between number of audiences' comments and important scenes for the audiences. A scene which has sufficient number of comments is appropriate as a part of the summarized video.*

We assume audiences' comments increases when it is an important scene because the comments would represent audiences' willingness to watch the scene. The scenes which have a lot of comments would be worth watching for the other audiences and would be also important part of the summarized video. If the hypothesis is correct, we can get a set of candidate scenes for video summarization and generates the summarized video by putting several candidate scenes together.

Expected issues are that the highly-commented scenes are just interesting scenes for the audiences and they are not parts of the summarized scenes. In this case, the scenes are

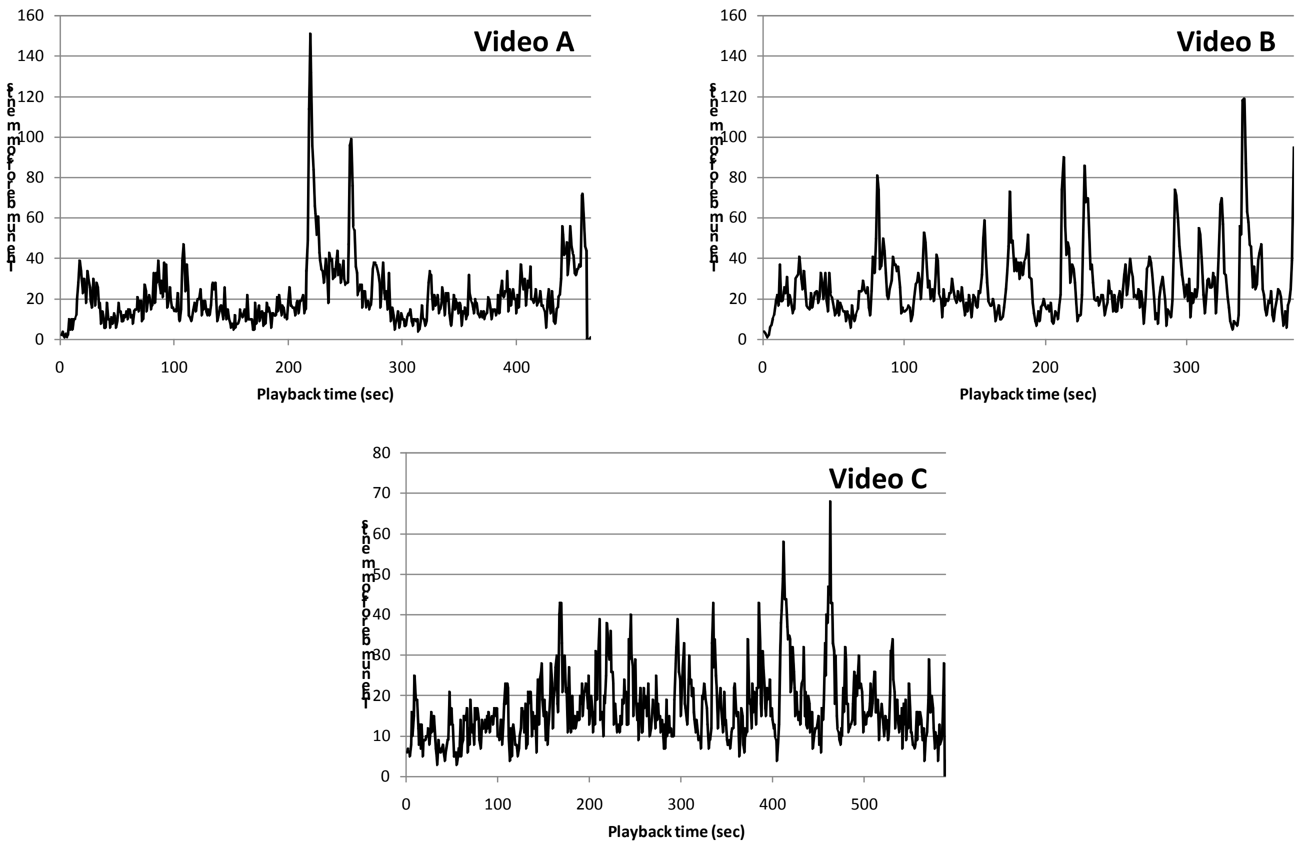

Figure 2: Changes in the number of comments per second.

a set of candidates for highlight. We need to study the relationship between the number of comments and summarized/interesting scenes.

4 EXPERIMENTAL ANALYSIS

We conducted an experimental analysis to verify our hypothesis. For the analysis, we collected audiences' comments from a video sharing service and asked people to select scenes which are appreciate for summarized/interesting scenes. Then, we studied if the number of comments was positively correlated with summarized/interesting scenes.

4.1 Comment Collection

The comments data in the Nico Nico Douga is stored in a log database with the following information.

- Time and date when audiences commented
- Playback time when audiences commented
- User ID
- Comment
- Command to decorate the comment

We chose three popular videos (Video A, B and C) in Nico Nico Douga at random and collected ten thousand comments per video. The contents of the videos are as follows:

- **Video A**: A man makes a strange cake using a lot of cheep sweets and eats it. (Total length: 465 seconds)
- **Video B**: A man makes big balls of chocolate using a lot of small various chocolates and packages them. (Total length: 376 seconds)
- **Video C**: A man mixes various energy drinks and tries to drink the mixed one. (Total length: 589 seconds)

Each video has a story (introduction, making and completion). Figure 2 shows the changes in the number of comments per second. From the graph, high and low peaks can be clearly shown in each video. We presume these videos are suitable to verify our hypothesis and use them in the analysis.

4.2 Scene Selection

We asked 20 participants who are students in our university about the following questionnaire after watching each video. (Note: The order of watching the videos was at random for fairness)

1. Please select 5 scenes which are summarized the video on condition that each scene is 3 seconds.
2. Please select 5 scenes which are interesting in the video on condition that each scene is 3 seconds

After the questionnaire, we counted the selected times for summarized and interesting scenes. Figure 3 shows the results. We can see several differences between selected summarized scenes and interesting scenes in the results. In video A, there is an interesting scene around 400 seconds although it is not selected as a summarized scene.

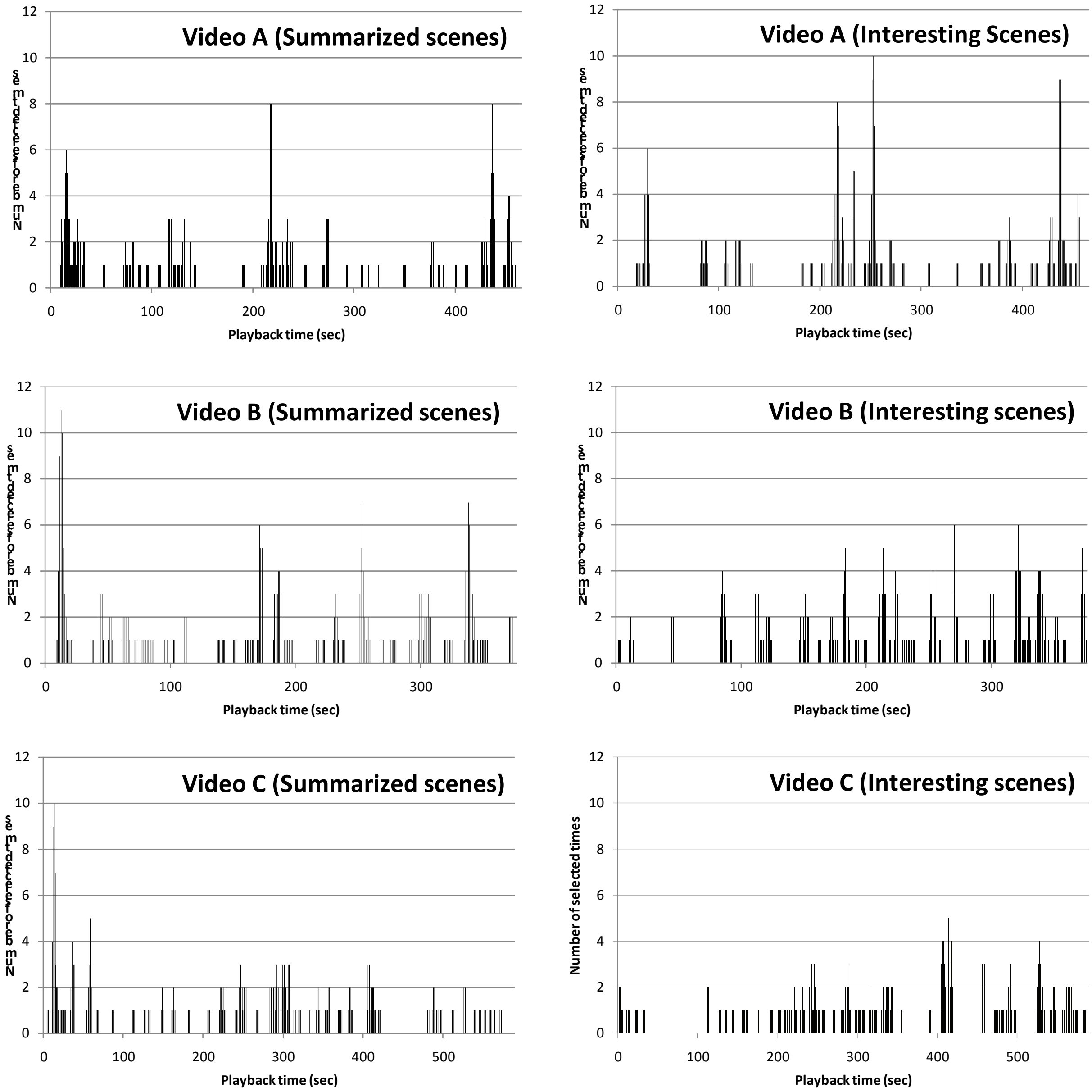

Figure 3: The number of selected times for summarized and interesting scenes.

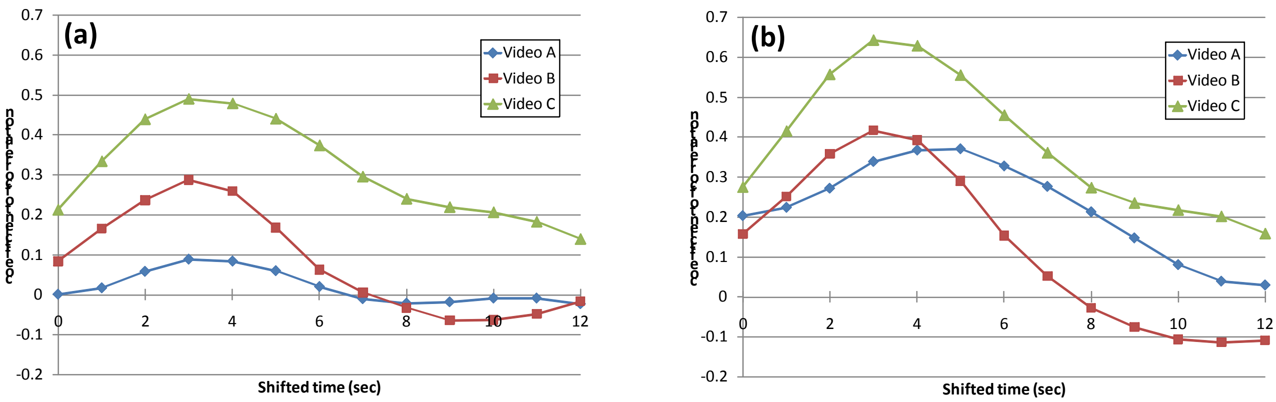

Figure 4: (a): coefficient of correlation between the number of comments and summarized scenes. (b): coefficient of correlation between the number of comments and interesting scenes.

The scene shows interesting performance but it is not important to explain the story of the video. In the video B, we can see a scene which is selected as a summarized scene at the beginning of the video although it is not interesting. Because the scene shows the title of the content, it is selected despite it is not interesting. The same thing can be also said for the video C. At the beginning of the video C, a man explains the purpose of the video. Of course, it is not so interesting but important for summarization. Thus, we can see summarized scenes do not always correspond with interesting scene and a scene of introduction is important for video summarization even if it is not interesting one.

4.3 Analysis

We assessed coefficient of correlation between the number of comments and the number of selected times for summarized/interesting scenes. The result, however, does not show correlation between them. We presume that audiences would need to type their keyboard for a few seconds to comment to a scene and the input time should be required. Therefore, we shift the commented time to a few seconds before and assessed the coefficient of correlation again.

Figure 4 shows the result when the commented time is shifted by 1 second. From the graph, we can see the number of comments was positively correlated with summarized/interesting scenes when commented time was shifted to from 3 to 5 seconds before in these 3 videos. In the video A, the coefficient of correlation between the number of comments and summarized scenes is 0.09 when shifted to 3 seconds and 0.37 when shifted to 5 seconds as for interesting scenes. Weak correlation is shown only between the number of comments and interesting scenes. In the video B, the coefficient of correlation between the number of comments and summarized scenes is 0.28 when shifted to 3 seconds and 0.42 when shifted to 3 seconds as for interesting scenes. Weak correlation is shown between the number of comments and summarized scenes, and medium correlation as for interesting scenes. In the video C, the coefficient of correlation between the comments and summarized scenes is 0.49 when shifted to 3 seconds and 0.64 when shifted to 3 seconds as for interesting scenes. Medium correlation is shown between the number of comments and summarized/interesting scenes.

We found the number of comments was positively correlated with summarized/interesting scenes when the commented time was appropriately modulated in consideration of input time. The correlation strength differs in the contents of videos and the coefficient of correlation of interesting scenes is higher than that of summarized scenes.

5 DISCUSSION

The experimental analysis clarified summarized scenes do not always correspond with interesting scenes and the coefficient of correlation of interesting scenes is higher than that of summarized scenes. There are two issues. The first issue is how to extract a scene which is important for video summarization but few comments. The second issue is how to exclude scenes which have a lot of comments but inappropriate for summarized scenes. To solve the issues, we have two main approaches. The first approach is to make a support system which extracts candidate scenes using audiences' comments and suggests the scenes to users so that they can make a summarized video quickly and improve its quality. In this approach, the users decide whether the suggested scenes are appropriate or not and find missing scenes. The advantage of the first approach is ease of implementation and the drawback is workload of the users. The second approach is to devise an algorithm which finds unnecessary and missing candidates. We presume the number of comments is not sufficient as a parameter for the algorithm and additional parameters are required. For the additional parameter, meaning of the comments would be effective. Moreover, we probably need to use audio-visual summarization techniques together. The advantage of the second approach is to reduce human workloads and the drawback is difficulty of implementation. Since each approach has different advantages, we will study the two approaches as future work.

Compared with existing summarization schemes, the proposed scheme could produce more appropriate summarized video in terms of audience-oriented aspect. Traditional audio-visual video summarization techniques can detect importance of the scenes in terms of audio-visual aspect but cannot understand context of the scenes. The audience comments can represent context of the scenes and it can be regarded as metadata of the video which is described by the audience. Although there are some researches which use metadata of a video described by its producers for video summarization [16], we presume the proposed scheme could realize more audience-oriented video summarization because it uses metadata of the video described by themselves.

The experimental analysis also clarified commented time should be shifted to several seconds because of input time for comment messages. However, accurate time of the gap is not clear yet and we should estimate the gap time. One of the solutions is to focus on length of the comments and estimate the input time by multiplying average time for inputting one character by the length. The average input time would vary from person to person but it would be able to approximate the input time. In this case, we would have to take into account the combination of the inputted characters in order to estimate the input time more accurately.

Although we use collected ten thousand comments for the analysis in the experiment, the minimum number of comments required for extraction of summarized scenes should be discussed. Since there is no comment when a user uploads a video to a video sharing site, our proposed scheme cannot be applied and only audio-visual summarization techniques are effective. As time passes, audiences' comments are collected and our proposed scheme can be applied. By combining audience-driven summarization with audio-visual summarization, we presume the videos can be summarized more appropriately for audiences because it is difficult to know meaning of the scenes and audiences' interests if there is only audio-visual information. We should study the threshold of number of comments to apply the audience-oriented video summarization by changing the number of comments.

6 CONCLUSION

In this paper, we proposed an audience-oriented video summarization scheme which analyzes audiences' feedbacks in the video and finds important scenes for video summarization in audiences' point of view. From the experimental analysis using audiences' comments in Nico Nico Douga, we got five findings; (1) summarized scenes do not always correspond with interesting scene, (2) a scene of introduction is important for video summarization even if it is not interesting one, (3) there is a short-time delay between comments and target scene, (4) the number of comments was positively correlated with summarized/interesting scenes when commented time was shifted to from 3 to 5 seconds before, (5) Some schemes would be required to make summarized video from audiences' comments because the audiences' comments indicated interesting scenes rather than summarized scenes.

As future work, we will design a support system for video summarization while studying an algorithm of video summarization based on the meaning of the comments so that we can generate summarized videos automatically. We will also compare the audience-driven video summarization method with some audio-visual summarization methods in order to show effectiveness of the proposed method more clearly.

REFERENCES

[1] YouTube, http://www.youtube.com/.

[2] Yahoo! Video, http://video.yahoo.com/.

[3] D. DeMenthon, V. Kobla, and D. Doermann, "Video summarization by curve simplification", ACM Multimedia, pp. 211-218 (1998).

[4] Y. Ma, L. Lu, H. Zhang, and M. Li, "A user attention model for video summarization," ACM Multimedia, pp. 533-542 (2002).

[5] C. Kim and J.N. Hwang, "An integrated scheme for object-based video abstraction", ACM Multimedia, pp. 303-311, 2000.

[6] B. T. Truong and S. Venkatesh, "Video abstraction: A systematic review and classification," TOMCCAP, Vol. 3, Issue 1 (2007).

[7] H. Luo, Y. Gao, X. Xue, J. Peng, and J. Fan, "Incorporating Feature Hierarchy and Boosting to Achieve More Effective Classifier Training and Concept-Oriented Video Summarization and Skimming," TOMCCAP, Vol. 4 Issue 1 (2008).

[8] X. Tong, Q. Liu, Y. Zhang, and H. Lu, "Highlight Ranking for Sports Video Browsing," ACM MULTIMEDIA, pp. 519-522 (2005).

[9] M. Fleischman, B. Roy, and D. Roy, "Temporal feature induction for baseball highlight classification," ACM MULTIMEDIA, pp. 333-336 (2007).

[10] B. Yu, W.Y. Ma, K. Nahrstedt, and H.J. Zhang, "Video summarization based on user log enhanced link analysis," ACM Multimedia, pp. 382-391 (2003).

[11] Y. Rui, A. Gupta, and A. Acero, "Automatically extracting highlights for TV Baseball programs," ACM Multimedia, pp. 105-115 (2000).

[12] D.A. Shamma and R. Shaw, "Watch what I watch: using community activity to understand content," Proc. of the international workshop on multimedia information retrieval, pp. 277-284 (2007).

[13] Ustream.tv, http://www.ustream.tv/.

[14] Stickam, http://www.stickam.com/.

[15] Nico Nico Douga, http://www.nicovideo.jp/.

[16] Y. Takahashi, N. Nitta, and N. Babaguchi, "Video Summarization for Large Sports Video Archives," IEEE International Conference on Multimedia and Expo, pp. 1170-1173 (2005).

(Received August 28, 2009)
(Revised November 15, 2010)

Yoshia Saito received his Ph.D. degree from Shizuoka University, Japan, in 2006. He had been an expert researcher of National Institute of Information and Communications Technology (NICT) from 2004 to 2007, Yokosuka, Japan. He is currently a lecturer at Iwate Prefectural University since October 2007. His research interests include internet broadcasting and interactive TV. He is a member of IEICE, IPSJ, IEEE, and ACM.

Yoshiki Isogai received his bachelor's degree from Iwate Prefectural University, Japan, in 2009. He currently attends a graduate course in Iwate Prefectural University. His research interests include Internet broadcasting and interactive TV. He is a member of IPSJ.

Yuko Murayama is a professor at Iwate Prefectural University.
She had M.Sc. and Ph.D. both from University of London in 1984 and 1992 respectively. She had been a visiting lecturer from 1992 to 1994 at Keio University, a lecturer at Hiroshima City University from 1994 to 1998. She has been with Iwate Prefectural University since April 1998. Her interests include internetworking, network security and trust. She is a member of IEEE, ACM, IPSJ, IEICE, and ITE. Currently she serves IFIP TC11 as a Vice Chair.