Personal TV Viewing by Using Live Chat as Metadata

Hisashi Miyamori
National Institute of Information and Communications Technology (NICT)
3-5 Hikari-dai Seika-cho Souraku-gun, Kyoto, 619-0289 Japan
+81 774 98 6873
miya@nict.go.jp

Satoshi Nakamura
National Institute of Information and Communications Technology (NICT)
3-5 Hikari-dai Seika-cho Souraku-gun, Kyoto, 619-0289 Japan
+81 774 98 6873
gon@nict.go.jp

Katsumi Tanaka
National Institute of Information and Communications Technology (NICT)
Kyoto University
Yoshida Honmachi, Sakyo, Kyoto, 606-8501 Japan
ktanaka@i.kyoto-u.ac.jp

ABSTRACT
We propose a new TV viewing method by personalizing TV programs with live chat information on the Web. It enables a new way of viewing TV content from different perspectives reflecting viewers' viewpoints.

Categories and Subject Descriptors
H.2.4 [**DATABASE MANAGEMENT**]: Systems – *multimedia databases,* H.5.1 [**INFORMATION INTERFACES AND PRESENTATION**]: Multimedia Information Systems – *video*.

General Terms: Algorithms, Management, Documentation, Human Factors

Keywords: live chat, fusion of broadcast and web content, metadata generation, viewer, viewpoint, semantic analysis, digest

1. INTRODUCTION
Recently, the amount of TV program data that can be recorded at home has been rapidly increasing with the improvement in the performance of hard-disk (HD) recorders. Currently, 600GB-capacity HD recorders have appeared in the market with their recording capability of more than 1070 hours at a certain quality. Because users do not have an unlimited amount of time to view such content, however, there is a great need for functions that can efficiently search only necessary segments of video from a huge amount of recorded data, present an overview of the content in a compact form, or provide a digest in a limited amount of time.

Video indexing is one of the fundamental techniques to achieve such functions. Conventionally, several multimodal indexing methods have been proposed using visual features such as color[1], camera motion[2], human faces[3], texts obtained from closed captions[4], classes and volumes of audio information[5]. However, because these methods are based on data provided by broadcast stations, the obtained indices basically reflect only the intentions of TV programmers and stations. Therefore, conventional methods cannot incorporate factors such as the responses of other viewers of a TV program into functions of scene search, summary presentation, and digest viewing.

Meanwhile, growing attention has been recently focused on live chat on the Web. Live chat communities on the Internet are virtual communities where viewers of a TV program congregate, post messages in real time about their impressions or the program itself, and have fun with such conversations. By enjoying the chats in parallel with a TV program, users can feel a sense of unity by virtually sharing emotions with other viewers.

We propose a new TV viewing method by personalizing TV programs with live chat information on the Web. Statistical analysis and pattern recognition of the data on the chat can effectively extract metadata related to the viewer's viewpoint such as important scenes in the program or responses by a particular viewer. The extracted metadata is expected to enable a new way of viewing TV content from different perspectives reflecting viewers' viewpoints.

2. OVERVIEW OF PROPOSED METHOD
The overview of the proposed method is shown in figure 1.

Although it depends on settings, a piece of information exchanged on a live chat is basically composed of three components: the time when the message was posted, the ID of the person who posted the message, and the content of the message, as shown in figure 2.

First, the recorded chat data are parsed to extract the time when the message was posted, the ID of a person who posted the message, and the content of the message. The post time is often out of sync with the time when the corresponding event happened in the program. This is because the viewer posts messages after events happen in the video. In this paper, the post time is adjusted by calculating the compensation time corresponding to the time it takes to create the text of the posted message.

Then, the intensity of responses is calculated by counting the number of message entries in a unit of time. Also, the emotional intensities of viewers' enjoyment and depression are extracted by detecting the frequencies of ASCII art expressions that appear in the messages by pattern matching. Practically, the detection can be improved by incorporating a learning approach using words and phrases representing enjoyment such as: “good,” “great,” “amazing,” “wow,” and “gee,” or words and phrases representing depression such as: “alas,” “no,” “uh-oh,” “boo-hoo”, and “sigh”.

As a result, the following group of indices, composed of sequences of values, is generated. For total viewers: the measured time, the measured unit of time, the intensity of responses, and the intensity of enjoyment/depression. For individual viewers: the ID of a person who posted the message, the measured time, the measured unit of time, the intensity of responses, and the intensity of enjoyment/depression.

WWW 2005, May 10-14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

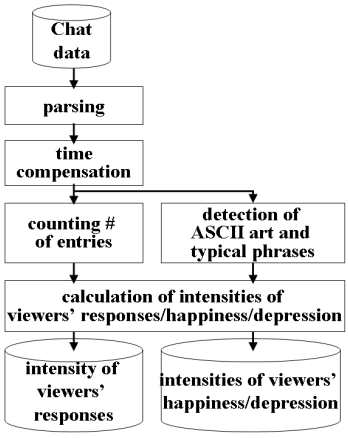

Figure 1. Processing overview of the proposed method

80 Name: anonymous 04/10/19 08:48:22 ID:R9C+YqFS
What a severe blitz, but here it comes!
Touch——————(ﾟ∀ﾟ)——————Down!!

81 Name: anonymous 04/10/19 08:48:28 ID:+FNfUHkk
Unstoppable. They might get more than 50 points…

82 Name: anonymous 04/10/19 08:48:30 ID:BZZg6/Cd
Only in the 1st quarter…

83 Name: anonymous 04/10/19 08:48:48 ID:7iYg967m
Player A is easily offended at the time like this.
His jumping pass is unforgivable considering his height.

84 Name: anonymous 04/10/19 08:49:15 ID:3bLzDEmh
Dogfight might be easier for beginners to understand.

85 Name: anonymous 04/10/19 08:49:23 ID:R9C+YqFS
Many blitz. Wonder how many were left in the back…

Figure 2. Conversation in a live chat

3. IMPLEMENTATION OF PROTOTYPE

An implementation example of a TV viewing screen using the metadata generated by the proposed method is illustrated in figure 3. In the figure, the headings are displayed using the closed captions segmented into topics, and the partial video corresponding to each topic is shown as a thumbnail. Hyperlink texts on the right of each thumbnail are also displayed using the closed captions. Below that, texts indicate the viewers' responses or the degree of emotion with facial icons representing emotions and some message texts obtained from the live chat.

One of the biggest advantages of the proposed method using a live chat is that the method can incorporate factors such as the viewpoints and responses of other viewers of a TV program into the generated metadata. The proposed method introduces a new factor of viewers' responses, which did not exist in the conventional methods, as shown in figure 4. The proposed method provides viewing methods with various added values such as: chronological/ranking display using the viewers' responses, viewing that shares the viewers' intensities of responses, enjoyment, or depression, discovery or sharing of the sense of value of other viewers with similar/different interests as the user.

The implemented prototype demonstrated that the proposed method based on the time when the message was posted, the ID of the person who posted the message, and the content of the message obtained from the live chat enables a new way of viewing TV programs from different perspectives. In addition, we expect to achieve a greater diversity of TV viewing if an infrastructure is available that actively distributes, manages, and utilizes viewer's information describing gender, age, district, and nationality, etc. For example, it is anticipated that one will be able to view TV in a way that reflects the viewpoint of men or women, of different generations, of districts or of nations. If users accept these viewing styles, various live chat communities will be created other than the so-called 2ch, which is one of the biggest Japanese bulletin board, in the future. However, there are many issues in this topic, including protection of private information.

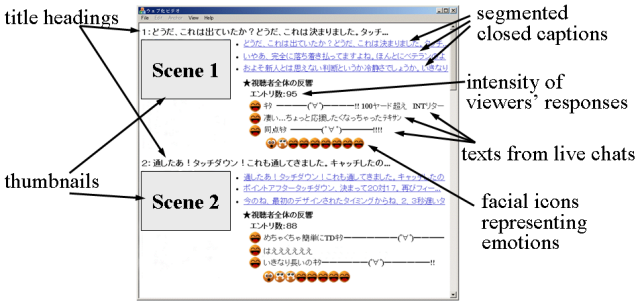

Figure 3. Implemented TV viewing screen using metadata generated by the proposed method

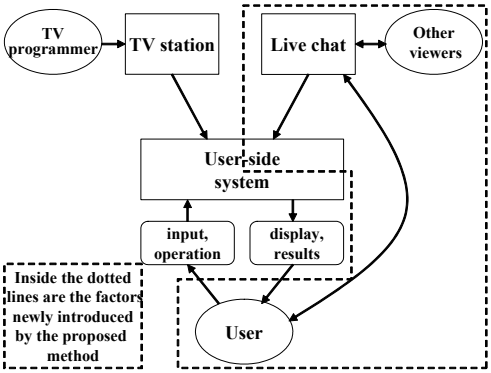

Figure 4. Concept of proposed method

4. REFERENCES

[1] Nagasaka, A., and Tanaka, Y., Automatic video indexing and full-video search for object appearances, IPSJ, Vol.33, No.4, pp.543-550, 1992.

[2] Akutsu, A., Tonomura, Y., Hashimoto, H., and Ohba, Y., Video indexing using motion vectors, SPIE Proc. VCIP '92, pp.522-530, 1992.

[3] Smith, M., and Kanade, T., Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques, CVPR, 1997.

[4] Nakamura, Y., and Kanade, T., Semantic analysis for video contents extraction - spotting by association in news video, ACM Multimedia, pp.393-401, 1997.

[5] Miyamori, H., Automatic annotation of tennis action for content-based retrieval by integrated audio and visual information, CIVR2003, LNCS2728, Springer-Verlag, pp.331-341, 2003.