Video Search by Impression Extracted from Social Annotation

Satoshi Nakamura and Katsumi Tanaka

Graduate School of Kyoto University
Yoshida-Honmachi, Sakyo, Kyoto, Japan
{nakamura,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract. This paper proposes a novel indexing and ranking method for video clips on video sharing Web sites that overcomes some of the problems with conventional systems. These problems include the difficulty of finding target video clips by the emotional impression they make, such as level of happiness, level of sadness, and so on because text summaries of video clips on video sharing Web sites usually do not contain such information. Our system extracts this type of information from comments on the video clips and generates an impression index for searching and ranking. In this work, we present analytical studies of video sharing Web site. Then, we propose an impression ranking method and show the usefulness of this method on the experimental test. In addition, we describe the future direction of this work.

Keywords: Video IR, ranking, indexing, impression, social annotation, video sharing Web sites.

1 Introduction

The popularity of video sharing Web sites has exploded over the past couple of years. *YouTube*, the main video sharing Web site in the world, had more than 80 million videos as of May 2008. *NicoNico Douga*, the main video sharing Web site in Japan, had about 2.7 million videos at the end of June 2009. On these sites, a vast number of users enjoy watching video clips. For example, *Ellacoya Networks* reported that nearly 79 million users watched more than 3 billion video clips on *YouTube* in January 2008 alone.

Video sharing Web sites have two types of users: uploaders who upload clips to the sites and viewers who view the uploaded clips. The basic procedure is that an uploader uploads a video clip with a title and a short summary. Then, the uploader and viewers add tags to the video clip to categorize it. A viewer can watch popular video clips by checking the video clip rankings and can find a target video clip by navigating with tags or by searching with keywords. However, it is not easy to find target video clips because the text information for each clip is very short. Particularly, the text information for each clip usually does not contain information about the type of emotional impression that the video clip might make happy, sad and so on.

On the other hand, people sometimes want to search video clips by impression. For example, when a user wants to lighten his mood, he may look for a funny video clip.

G. Vossen, D.D.E. Long, and J.X. Yu (Eds.): WISE 2009, LNCS 5802, pp. 401–414, 2009.

When a user wants to cry from watching a video clip, he may look for a tear-jerker. In addition, a user may want to watch video clips on subjects such as, for example, “amazing football technique” or “how to cook delicious food”. However, it is too difficult to find such video clips because such impression information for each clip is sparse and conventional systems provide only popularity-based ranking mechanisms and do not provide such impression-based searching and ranking. As a result, users may be unable to find clips relevant to their desired impression.

On *YouTube* and *NicoNico Douga*, users can post comments about a video clip, evaluating it or recommending it to other users. For example, *NicoNico Douga* had about two billion comments for about 2.7 million video clips at the end of June 2009. In addition, *YouTube* and *NicoNico Douga* have an embedded video service for Web pages, while many blog services enable bloggers to easily embed such video clips into their blogs. Many bloggers thus embed video clips that they recommend to their readers. Nevertheless, *YouTube* and *NicoNico Douga* do not use such social annotation to improve their search services.

In this work, we focus on using social annotation such as Weblogs, social bookmarks, and comments to generate indexes of video clips (Fig. 1). For example, in comments and Weblogs referring to a video clip, there may be comments about the user’s impression of the clip such as their evaluation of it, whether they enjoyed it, or if it made them feel sad. Such information is very useful for generating an index of video clips on video sharing Web sites for the purposes of searching and ranking.

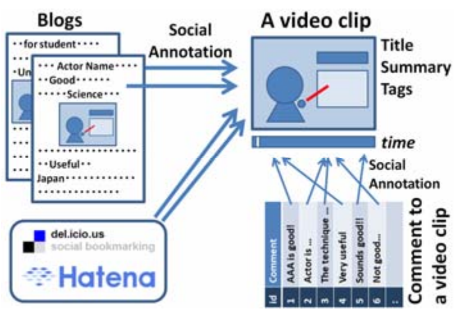

Fig. 1. Social annotation on video clips

Our ultimate goal is to develop a system for indexing and ranking video clips on video sharing Web sites that uses all relevant social annotation available on the World Wide Web such as comments on video sharing Web sites, comments on Internet bulletin boards, entries in Weblogs, tags in social bookmark services, and text on mash-up sites. As a first step, in this work, we propose a method for generating an impression index based only on comments about video clips. The impression indexing method enables users to search for or rank video clips based on feelings such as happiness, sadness, and surprise. We also developed a ranking algorithm based on the index and used a prototype system to experimentally evaluate our approach.

We first describe related work and explain the function of video sharing Web sites and social annotation. Then, we describe the results of our analytical study on the impact of social annotation on video sharing Web sites. Next, we describe our indexing and ranking methods and present the results of prototype testing, which shows the usefulness of our method. Finally, we conclude with a brief summary and a look at future work.

This work makes three significant contributions.

- It shows that comments are an important contribution to video sharing Web sites as a form of social annotation.
- It shows that impression indexing and ranking of video clips can be done by using comments found in social annotation.
- It experimentally shows that our impression index and its ranking are useful to look for target video clips depending on an emotional impression.

2 Related Work

Video indexing is a fundamental technique that enables users to search for a specific scene in a video clip or to generate a summary of a video clip. Several indexing methods have been proposed that use visual features such as color [2], camera motion [1], human faces [8], text obtained from closed captions [14], and classes and volumes of audio information [3, 4, 12]. These methods were mainly designed for use with broadcast TV programs, but they can be extended to video clips on video sharing Web sites. However, because they use only data provided by the content provider, the indexes generated basically reflect only the provider's intentions. These methods thus cannot incorporate factors such as the viewpoints and responses of viewers into the search and ranking functions for video clips.

Dimitrova et al. proposed a content-based video retrieval method that uses an example video clip [9]. The content-based approach is one method of retrieving video clips. We approach the video-retrieval problem differently. We will show the usefulness and potential of social annotation for video retrieval.

We proposed and developed a system for generating a summarizing video of a TV program by analyzing comments on an Internet bulletin board about the program [5]. This system classifies comments into the categories of delight and sorrow by pattern matching with a delight/sorrow dictionary. The system then generates an index for making a digest based on the level of delight or sorrow. *Uehara* et al. described a system for creating an attention graph from dialogues on an Internet bulletin board about a TV drama [6]. This system detects the level of viewer attention by analyzing the comments for each scene in the drama. These researches only focused on searching for specific scenes within a video clip and did not focus on searching for a video clip from a large video clip database. In addition, these researches did not address the generation of an index for searching for and ranking video clips.

Several methods have been described for using social annotation to judge the quality of content. For example, *Yanbe* et al. [15] and *Heyman* et al. [11] proposed using social bookmarks to rank Web search results. *Yanbe* et al. focused on using impression tags for Web pages to rank search results. *Boydell* et al. [10] proposed summarizing Web pages on the basis of social bookmarks. These efforts showed the potential of using social annotation for evaluating the quality of content. However, using social annotation to generate an index of video clips and to judge their quality has not been addressed.

3 Video Sharing Web Site

Millions of video clips have been uploaded to video sharing Web sites, and millions of users watch them. In the work reported here, we used *NicoNico Douga* as the video sharing site as it is the most popular video sharing Web site in Japan. It had about 10 million users as of the end of October 2008. Users can upload, view, and share video clips as they do on *YouTube* and other video sharing sites. The differences between *NicoNico Douga* and the others are the simplicity of posting comments at specific points in a video clip and a function that enables users to overlay posted comments on a video clip.

While *YouTube* users can also post comments for a video clip, it is not easy to post comments at a specific playback point. Instead, commenters include the target playback time in their posted comments such as “Watch him fall at 2:30!” When a *NicoNico Douga* user posts a comment for a video clip he is watching, the system sets the playback time of the video clip at the time the comment was posted as the target playback time of the comment. The user can easily post comments for a specific time point in a video clip with this system.

Moreover, *NicoNico Douga* overlays the comments for a video clip at the corresponding playback times. Users enjoy not only watching the video clip but also seeing the comments of others at the appropriate points in the video. This synchronicity creates a sense of a shared watching experience. This comment overlay function can be turned on and off by the user.

As mentioned, *NicoNico Douga* had more than 1.9 billion comments for 2.5 million video clips as of May 2009. There were more than 10 million comments for the most commented upon video clip! We can thus say that comments make an important contribution to video sharing Web sites as a form of social annotation. We will address their impact more specifically in the next section.

We believe that such video sharing sites will become even more popular worldwide, and that the number of video sharing sites will continue to increase. In fact, several video sharing sites have followed the lead of *NicoNico Douga* and have started providing synchronous comment services (*LYCOS mix*[1] in Japan, *AcFun*[2] in China, and so on). In addition, some mash-up sites have started to manage posted comments and overlay them on video clips that are stored on other video sharing Web sites. The alpha version of *NicoNico Douga* was also a mash-up site that used video clips stored on *YouTube*.

On *NicoNico Douga* and similar sites, the information and social annotation for each video clip usually include the following:

- Identification number of a video clip
- Title and summary of the video clip, which are written by the uploader
- Number of times viewed, number of posted comments, and number of times it has been marked as a favorite
- Upload date and length of the video clip
- Tags added by users
- Viewer comments.

[1] http://mix.lycos.jp/
[2] http://www.acfun.cn/

A viewer comment generally includes the identification number of the viewer, the comment itself, the date posted, and the corresponding playback time.

NicoNico Douga uses the number of views, comments, and favorite settings to rank video clips on the assumption that these metrics reflect popularity. While this may be sufficient in terms of determining overall popularity, it is insufficient for indicating the quality of a video clip. For example, it is not easy for users to search for tear-jerker video clips. In addition, they cannot rank video clips by the level of "tear-jerker-ness." Conventional systems do not provide such indexing or ranking systems. Our method for generating an impression index for video clips does.

4 Analytical Study

First, we created two sub-datasets of the weekly and monthly 100 most commented upon video clips to evaluate the usefulness of the number of comments about a clip in *NicoNico Douga*. Here, we manually assigned "low quality video clip" to video clips that are specifically focused on collecting comments (i.e., the uploader asks viewers to post comments), video clips that are typing games (i.e., viewers type text in response to presented text), and video clips made for greeting each other and so on. Figure 2 shows the relationship between the rankings of video clips based on the number of posted comments and the number of low quality videos. In this figure, the horizontal axis is the ranking based on the number of posted comments and the vertical axis is the number of low quality video clips.

We found that about 18% of the weekly top 100 clips and 32% of the monthly top 100 clips were low quality. In addition, there were more low quality video clips among those that ranked the highest than among the low ranked video clips based on the number of posted comments. This result indicates the number of comments is insufficient for judging the quality of video clips.

To construct a dataset for analyzing *NicoNico Douga*'s comments, we developed a comment crawler that collect the comments and some information such as the title,

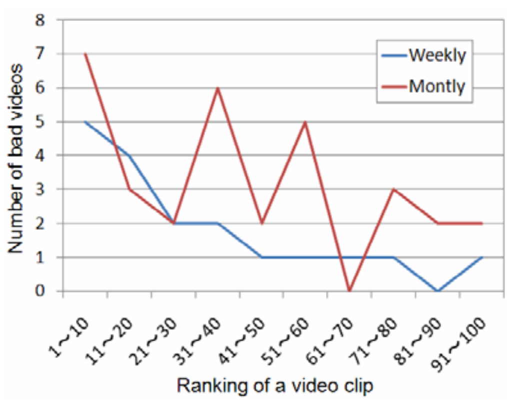

Fig. 2. The relationship between the rankings of video clips based on the number of posted comments and the number of low quality videos

summary, and tags and so on. Each *NicoNico Douga* video clip is identified by a unique number and the largest video clip identification number was just over 5,000,000 when we started to crawl them (October 21, 2008). Our crawler generates an identification number randomly from 1 to 5,000,000 to crawl them. We limited the crawling to the most recent 1,000 comments per clip because, as mentioned, a video clip can have up to 10 million comments.

We ended up with 968,721 video clips (19.4% of all clips on *NicoNico Douga*). Although the number was relatively small compared to the total number of clips, it was sufficient for analyzing the impact of social annotation.

We divided the dataset into live video clips, which users could watch, and dead video clips, which users could no longer watch because they had been deleted. There were 304,460 live video clips and 664,261 dead video clips. This means that 68.57% of the video clips in our dataset had been removed either because the uploaders had removed them or because the service had removed them due to copyright violations. On this site, copyright violation is a major reason for removal. The number of crawled comments for the live video clips in our set was 56,473,136.

The video clips in our dataset had an average length of 549.44 seconds. Moreover,

- The average number of viewings was 4072.73.
- The average number of comments was 479.88.
- The average number of times a video clip was marked as a favorite was 56.67.

Figure 3 shows the relationship between the number of posted comments and the percentage of the video clips in our dataset that had that number of comments. In this figure, the horizontal axis is the percentage of video clips in our dataset and the vertical axis is the number of posted comments per video. As shown by the plot in this figure, about 38.5% of the video clips in our dataset had more than 100 comments, and about 5% had more than 2000.

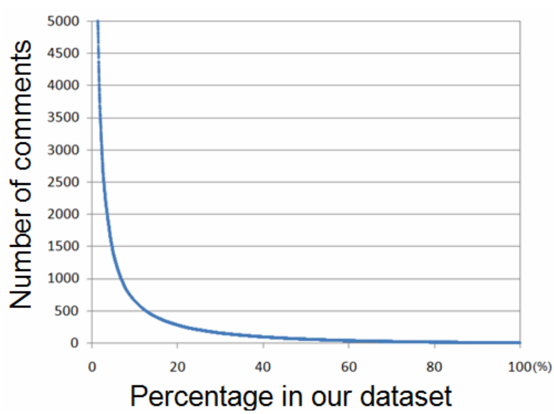

Fig. 3. Percentage of video clips with specified number of posted comments

We developed a dictionary that supported the generation of an impression index to classify comments as either positive or negative, and as indicating happiness, sadness, and surprise [16].

In this dictionary, there are 217 patterns of regular expressions to match positive comments, 232 patterns to match negative comments, 13 patterns of regular expressions to match comments expressing happiness, 30 patterns to match comments expressing sadness, and 7 patterns to match comments expressing surprise. We generated these regular expressions manually to detect the type of impression of the comments.

Here, we randomly selected 10,000 comments from our dataset to check the accuracy and coverage of extracting each factor. “Accuracy” is the percentage of extracted comments that are correct, i.e., they match the target impression. “Coverage” is the percentage of correct comments that are extracted. The correct comments were manually identified.

$$Accuracy = \frac{Num(extracted\_correct\_comment)}{Num(extracted\_comment)} \times 100$$

$$Coverage = \frac{Num(extracted\_correct\_comment)}{Num(correct\_comment\_in\_dataset)} \times 100$$

As shown in Table 1, the accuracy was a little lower for “happiness” than for other impressions. The reason for the low accuracy of detecting “happiness” comments was that viewers use the laughing symbol not only for laughing but also mockery. To solve this problem, we have to analyze comments in detail.

Table. 1. Accuracy and coverage

	Accuracy	Coverage
Positive	95.3%	97.2%
Negative	97.1%	93.7%
Happiness	85.5%	98.3%
Sadness	95.8%	97.5%

If the number of comments for a video clip is small, our system processes have lower reliability. However, 38% of video clips have more than 100 comments. We can say that their accuracy and coverage are sufficient to judge the level of impressions or to rank video clips according to impression.

We then used our dictionary to judge the impression of each comment in our dataset. On average, for each video clip, there were 22.24 positive comments, 10.25 negative comments, 71.24 comments expressing happiness, 6.32 comments expressing sadness, and 2.18 comments expressing surprise. That is, there were relatively more comments that were positive or that expressed happiness. We can say that there are many positive comments and comments expressing happiness and there are few comments expressing surprise or sadness. These average values are useful for judging the types of the impression of a video clip.

Next, to analyze the relationship between the impression comments and their corresponding playback times, we created a sub-dataset containing those video clips in the original dataset with more than 100 comments. This sub-dataset contained

117,217 video clips. In this analysis, first, our system normalizes the playback time of the video clip by dividing it into 100 units of playback time. Then, the system calculates the ratio of total comments and the ratio of each type of impression comment in each playback unit and each video clip. Finally, the system calculates the average of these in each playback unit.

Figure 4 shows the change in the number of comments by impression as video viewing progressed. The horizontal axis represents the video playback time in percentage terms. The vertical axis represents the ratio of comments for each impression.

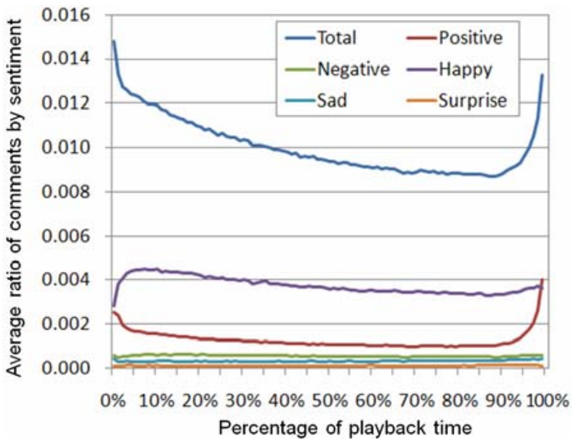

Fig. 4. Average ratio of comments by impression as video viewing processed

We found that the ratio of comments that were negative, sad, or expressed surprise, on average, was fairly evenly distributed over the playback time. In contrast, the average ratio of total comments decreased from the start to about 90% of the playback time and then sharply increased. Moreover, the ratio of positive comments decreased slightly from the start to about 90% of the playback time and then also sharply increased until the end of viewing; the ratio of positive comments at the end was twice that at the start. We can use these results to normalize the level of each impression at each point in time during playback or as a threshold for assigning one or more impressions to a video clip.

Here, we extracted 1519 enjoyable video clips that were tagged “enjoyable” or “laughter,” and 560 tear-jerker video clips which were tagged “tear-jerker” or “moving” from our dataset by matching the tags. Our system normalized video clips with 20 playback units.

Figures 5 and 6 show the change in the number of comments by impression as video viewing progressed. Figure 5 relates to enjoyable video clips and Figure 6 relates to tear-jerker video clips. In these figures, the horizontal axis represents the video playback time in percentage terms and the vertical axis represents the ratio of comments for impression.

We found that tear-jerker video clips had many comments expressing sadness, more than enjoyable video clips, and more than the average of all video clips in our

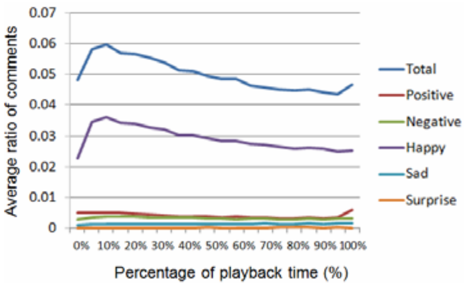

Fig. 5. Average ratio of comments by impression as video viewing processed in enjoyable video clips

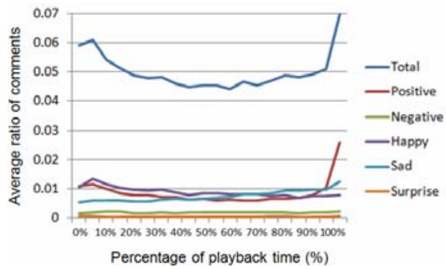

Fig. 6. Average ratio of comments by impression as video viewing processed in tear-jerker video clips

dataset. In addition, we also found that the end of tear-jerker video clips had many positive comments, more than the end of enjoyable clips. We can use these differences to determine the type of video clip.

5 Our Method

5.1 Impression Indexing and Ranking

Using the results of our analyses, we developed a method for ranking video clips for impression searching.

Our system uses this method not only for ranking video clips but also for searching by impression.

Our system first normalizes the playback time of a video clip by dividing it into 100 units of playback time. Next, the system counts each type of impression comment in each playback unit using our dictionary. Then, the system calculates the impression score of the video clip using the following equation.

$$Score(v,i) = \log_{10} impression_{all} \times \frac{impression_{all}}{total_{all}} + w(s) \times \frac{\sum_{k=96}^{100} positive_k}{total_{all}}$$

where v is the target video clip, s is the target impression the user searches for $impression_{\text{all}}$ is the total number of comments expressing target impression s, $total_{\text{all}}$ is the total number of comments to the video clip, $w(i)$ is a weight value for target impression i, and $positive_{\text{k}}$ is the number of positive comments in the k^{th} playback unit. In this equation, we emphasize the positive comments at the end of the video clip based on the results shown in Fig. 4. In addition, we set w(sadness) as higher than w(happiness) because of the results shown in Figs. 5 and 6.

When the user searches for an impression i for a list of extracted video clips, the system calculates the score of each impression. The clips are sorted on the basis of the scores.

If a query contains an impression keyword (i.e., moving, tear-jerker, laughter, happiness, sadness, surprising, and so on) that is defined in the query modification dictionary we prepared, our system uses these terms not only for the keyword search but also for impression-based ranking.

When a user submits an impression term with other keywords as a query, our simple query modification mechanism first extracts video clips with the other keywords and then sorts by the level of the input impression.

For example, when the user inputs “tear-jerker cat story” as a query, the system extracts video clips that contain “cat story”, sorts them by the level of sadness, and displays them.

5.2 Implementation

We developed our crawling system using Perl. The system crawls the video clip comments and the title, summary, tags, posted date, length, and so on and stores them in a database.

When the user inputs a query that does not contain an impression term, the system first returns a ranking of video clips based on *NicoNico Douga*’s popularity-based rankings. Then, our system enables users to rank the list of video clips on the basis of happiness, sadness, surprise, positive response, and negative response by clicking the impression button. After that, the system re-ranks the list of video clips based on the calculated impression score.

In addition, we also developed our client system as an extension of Mozilla Firefox 3.0. This system automatically generates a time-related graph for each video clip when the user accesses a ranking page showing the video search results or a video clip page. Figure 7 compares the conventional system and our system. The conventional system has no function to rerank the search results and only provides a thumbnail image, posted date, title, summary, recently posted comments, length, and some other information.

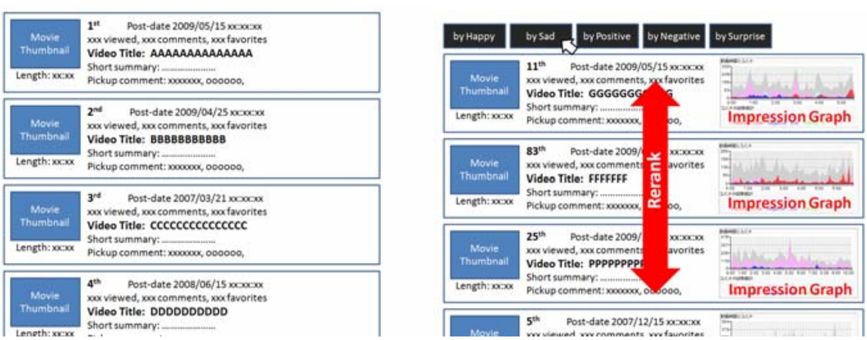

Fig. 7. The left figure is an image of a list of search results using the conventional system. The right figure is an image of a list of search results using our system. Our system shows impression graphs and has several control buttons the user can click to rerank the search results.

Figure 8 is a screen snapshot of our system. With this system, a user can easily see how the impression levels changed during viewing, enabling him or her to judge the quality of a video clip before watching it. Figure 9 shows an example of an impression graph. Figure 10 shows a screen snapshot after reranking by level of sadness. The user can use our system without stress because our system can rerank 100 search results in only two seconds.

Fig. 8. Sample screen snapshot showing ranking of cooking video clips. Change in impression levels during viewing is shown on the right.

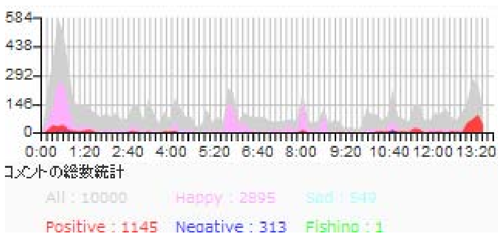

Fig. 9. An impression graph. The horizontal axis is the playback time. The vertical axis is the number of comments.

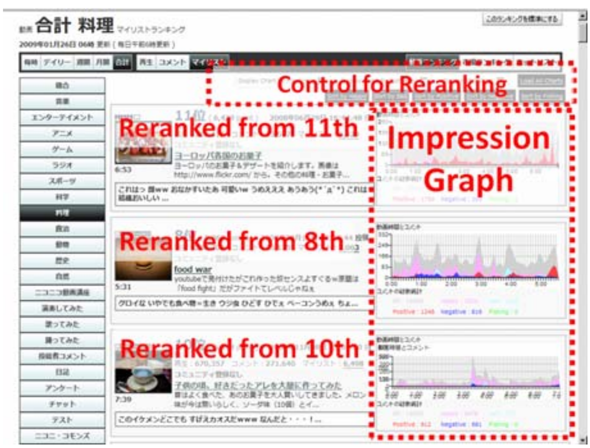

Fig. 10. Sample screen snapshot showing reranking of cooking video clips by level of sadness.

5.3 Evaluation

We evaluated our method experimentally to determine its usefulness to our system. In our evaluation, we used our collected dataset, which has 304,460 clips. In this experimental test, we conducted a user-based experiment to judge the usefulness of impression ranking.

We prepared five lists of enjoyable video clips and five lists of tear-jerker video clips, as determined by their tags. Each tag list of video clips had more than 10 video clips. We selected the top 10 most commented upon video clips in each list as a dataset. Then, we asked two users to judge the level of happiness and sadness of these video clips after watching them. They evaluated the video clips from 1 to 5. We did not inform the users of how the video clips had been ranked.

After collecting the user evaluations, we used our system to rank them according to the levels of happiness and sadness. Figure 11 plots these results. In this figure, the horizontal axis is the video clip rank based on our system and the vertical axis is the user evaluations of the video clips.

The results suggested that our system is useful for ranking video clips based on the level of sadness, but not for ranking clips based on the level of enjoyment.

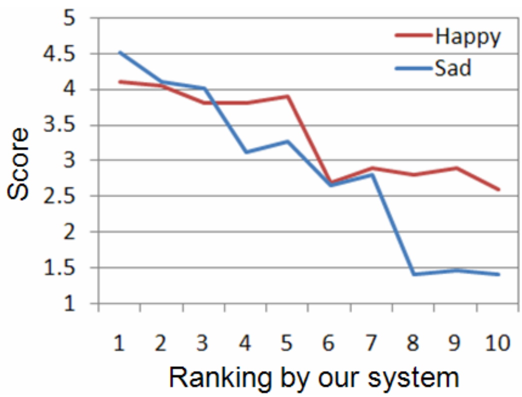

Fig. 11. Results of the impression ranking

6 Discussion

Our analytical studies showed the impact of comments as social annotation to video clips. They also showed how impressions such as happiness, sadness, and surprise change during viewing. We are confident that these results will be useful for future research in this area. Searching and ranking by impression will be one type of next-generation search system.

In a conventional system, the user cannot search for video clips based on impression information if the video clips do not contain any text about the impression, and, in fact, almost no video clips contain such impression information. Our system enables users to conduct an impression search. For example, a user can search for tear-jerker video clips, enjoyable video clips, and amazing video clips using our system.

There are many video clips that receive comments classified as indicating sadness. We asked our students to use our system and we received their feedback. We then found that high quality tear-jerker video clips had many comments expressing sadness throughout and many positive comments at the end. This knowledge supports our equation for calculating a sadness score.

Here, we focused only on happiness, sadness, surprise, positive response, and negative response. We can improve our system and develop new methods for ranking based on impression. We plan to improve our dictionary to detect additional impressions and context such as positive comments saying “thank you” and positive comments evaluating the video. If we can better utilize such comments we can improve our ranking method.

In addition, we plan to detect the senses related to posted comments (i.e., taste, sight, smell, touch, and hearing). If we can rank video clips on the basis of senses, the user can easily find video clips appealing to the sense of taste, beautiful video clips, video clips agreeable to the ear, and so on.

We did not take users or user groups into consideration. For example, users who support the *F.C. Barcelona* football team may enjoy video clips of matches lost by the *Real Madrid* football team or clips that viewers who support *Real Madrid* found disappointing. This impression is based on rivalry. There are many such situations. We thus plan to introduce group-based video ranking/recommendation.

We think that we can use posted comments to generate text indexes. For example, users post an actor name or event name to the specific playback time. Then, we can detect what happened or who acted and so on by analyzing posted comments. In addition, there have been many studies on detecting actors’ actions [7] or faces by image-based retrieval [13] methods. If we can combine these methods and our impression indexing method, we can create better indexing methods of video information retrieval.

8 Conclusion

In this paper, we showed the potential of video clip comments on video sharing Web sites and proposed an indexing and ranking method for searching video clips based on emotional impression extracted from these comments. Our impression indexing and ranking methods showed the potential of our system to contribute to next-generation video search techniques.

We did not consider blogs because the size of our crawled dataset was not large. We are now crawling blog entries to generate an index of video clips, and we plan to add this to our system and evaluate its usefulness. We think that the quality of content in blog entries is better than in video comments. Once we have introduced the use of blog entries for generating an index of video clips, we will focus on the differences in quality and quantity.

Acknowledgements

This work was supported in part by the National Institute of Information and Communications Technology and by the "Informatics Education and Research Center for Knowledge-Circulating Society" (Project Leader: Katsumi Tanaka, MEXT Global COE Program, Kyoto University). This work was also supported in part by MEXT Grant-in-Aid "Contents Fusion and Seamless Search for Information Explosion" (Project Leader: Katsumi Tanaka, A01-00-02, Grant#: 18049041).

References

1. Akutsu, A., Tonomura, Y., Hashimoto, H., Ohba, Y.: Video Indexing Using Motion Vectors. In: SPIE Proc. VCIP 1992, pp. 522–530 (1992)
2. Nakagawa, A., Tanaka, Y.: Automatic Video Indexing and Full-video Search for Object Appearances. IPSJ 33(4), 543–550
3. Saur, D., Tan, Y.P., Kulkarni, S., Ramadge, P.: Automated Analysis and Annotation of Basketball Video, Databases V. SPIE, vol. 3022, pp. 167–187
4. Miyamori, H.: Automatic Annotation of Tennis Action for Content-based Retrieval by Integrated Audio and Visual information. In: Bakker, E.M., Lew, M., Huang, T.S., Sebe, N., Zhou, X.S. (eds.) CIVR 2003. LNCS, vol. 2728, pp. 331–341. Springer, Heidelberg (2003)
5. Miyamori, H., Nakamura, S., Tanaka, K.: Generation of Views of TV Content Using TV Viewers' Perspectives Expressed in Live Chats on the Web. In: Proceedings of ACM Multimedia 2005, November 2005, pp. 853–861 (2005)
6. Uehara, H., Yoshida, K.: Annotating TV Drama based on Viewer Dialogue - Analysis of Viewers' Attention Generated on an Internet Bulletin Board. In: 2005 Symposium on Applications and the Internet (SAINT 2005), pp. 334–340 (2005)
7. Laptev, I., Perez, P.: Retrieving actions in movies. In: Proc. of ICCV 2007, October 2007, pp. 1–8 (2007)
8. Smith, M., Kanabe, T.: Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques. In: Bakker, E.M., Lew, M., Huang, T.S., Sebe, N., Zhou, X.S. (eds.) CIVR 2003. LNCS, vol. 2728, pp. 331–341. Springer, Heidelberg (2003)
9. Dimitrova, N., Abdel Mottaleb, M.: Content-based Video Retrieval by Example Video Clip. In: Proc. of SPIE, vol. 3022, pp. 59–70 (1997)
10. Boydell, O., Smyth, B.: From social bookmarking to social summarization: an experiment in community-based summary generation. In: Proceedings of the 12th international conference on Intelligent User Interfaces, pp. 42–51
11. Heymann, P., Koutrika, G., Garcia-Molina, H.: Can social bookmarking improve web search? In: Proceedings of the international conference on Web search and Web Data Mining, pp. 195–206
12. Intille, S., Bobick, A.: Closed-world Tracking. In: Proceedings of the Fifth International Conference on Computer Vision, pp. 672–678
13. Sivic, J., Everingham, M., Zisserman, A.: Person spotting: video shot retrieval for face sets. In: Leow, W.-K., Lew, M., Chua, T.-S., Ma, W.-Y., Chaisorn, L., Bakker, E.M. (eds.) CIVR 2005. LNCS, vol. 3568, pp. 226–236. Springer, Heidelberg (2005)
14. Nakamura, Y., Kanabe, T.: Semantic Analysis for Video Contents Extraction Spotting by Association in News Video. In: ACM Multimedia, pp. 393–401
15. Yanbe, Y., Jatowt, A., Nakamura, S., Tanaka, K.: Can Social Bookmarking Enhance Search in the Web? In: Proc. of. JCDL 2007, pp. 107–116 (2007)
16. Nakamura, S., Shimizu, M., Tanaka, K.: Can Social Annotation Support Users in Evaluating the Trustworthiness of Video Clips? In: Proc. of WICOW 2008 (2008)