Recompilation of Broadcast Videos Based on Real-World Scenarios

Ichiro Ide

Graduate School of Information Science, Nagoya University,
1 Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
ide@is.nagoya-u.ac.jp

Abstract. In order to effectively make use of videos stored in a broadcast video archive, we have been working on their recompilation. In order to realize this, we take an approach that considers the videos in the archive as video materials, and recompiling them by considering various kinds of social media information as "scenarios". In this paper, we will introduce our works in news, sports, and cooking domains, that makes use of Wikipedia articles, demoscopic polls, twitter tweets, and cooking recipes in order to recompile video clips from corresponding TV shows.

1 Introduction

In recent years, following the increase in the capacity of storage devices, research on retrieval and browsing of videos stored in a large-scale broadcast video archive has become active. When considering the retrieval of video media compared to text and image media, there is a problem that it is more difficult to browse and conceive the retrieved information at a glance.

Considering this problem, we have been working on methods that do not simply provide to the users, a list of individual video clips in the archive as the retrieved results, but instead, provide a presentation of recompiled video clips according to "scenarios" obtained from the real world; various kinds of social information available on, mostly, but not limited to, the Web.

The rest of the paper is organized in three parts, where we will introduce our works based on the above approach in the domains of news, sports, and cooking.

2 Applications to News Contents

For news contents, we will introduce two different works on video recompilation that makes use of different kinds of social information as "scenarios".

2.1 Description of Wikipedia Articles with Videos

- Video: News show
- Scenario: Wikipedia articles on current topics

S. Li et al. (Eds.): MMM 2013, Part II, LNCS 7733, pp. 58–68, 2013.

Fig. 1. The "Videopedia" interface. The left side of the screen shows the original chronological explanation texts in Wikipedia, while the right side shows the video clip corresponding to the explanation text high-lighted in the left-side. Video clips of news stories preceding and succeeding the corresponding news story along the topic thread structure are also shown above and below. In addition, links to other corresponding Wikipedia articles are listed next to each video clip.

Wikipedia is an online encyclopedia that allows the general public to edit at any time, so it usually contains up-to-date information on various phenomena in the real world. Here, we considered each Wikipedia article as a dynamic scenario, and proposed a method that visually describes it with the help of a sequence of video clips from the news video archive.

In order to realize this, we proposed a framework that links video clips from news shows (news stories) to chronological explanations texts in Wikipedia articles on current topics, and developed an interface "Videopedia" (Fig. 1) to demonstrate the results. Details of the method could be found in [1].

Linking Video Clips (News Stories) to Wikipedia Articles. First, we need to link corresponding video clips to Wikipedia articles. Here, each video clip represents a news story.

In order to narrow-down the candidates, date expressions are extracted from the chronological explanation in a Wikipedia article, and only video clips broadcast on or around the date are analyzed.

Then, the most related video clip is selected by comparing the term frequencies in the Wikipedia article and the closed-caption of the selected video clips. This process is performed for each date expression extracted.

Interpolation Based on the Topic Thread Structure. Since the detailedness of the descriptions are different in news shows and Wikipedia articles, the above process is not sufficient to obtain links to most of the date expressions. In order to compensate for this problem, the topic thread structure proposed in [2] was used to interpolate the links obtained in the above process.

The precision and the recall of the links obtained from the above processes were 82.1% and 75.8%, respectively, in a manual evaluation of three Wikipedia articles.

2.2 Generation of a Summarized Video on a Specific Person in News

- Video: News show
- Scenario: Demoscopic polls

In broadcasting stations, there are often cases that they need to produce a video that introduces a person's personal history. In most cases, such a necessity arises suddenly before the news show, so the producers need to gather source material and compile them in a limited period of time.

To provide an automatic solution for such a task, we focused on a "Prime Minister" as an example, since he appears in TV news quite frequently, and proposed a method to provide a biased summarized video that explains why he had to resign in the end. In order to produce such an explanation, we collected video clips corresponding to major events that occurred while he was in office, by referring to demoscopic polls and also features obtained from topic thread structures. In order to detect major events, we prepared two approaches: Template-based, which detects typical events for all Prime Ministers, and Topic-based, which detects major events specific to the period. Details of the method could be found in [3].

Preparation of Source Video Clips (News Stories). First, as the initial dataset, news stories that contain the Prime Minister's name as a subject in the closed-caption during his period in office are extracted.

Template-Based News Story Selection. In the inauguration / resignation periods (i.e. the beginning and the end of his period in office), there are typical events such as inauguration / resignation speeches, visits to foreign countries to see foreign leaders, and so on. In order to detect such stories, we prepared templates composed of typical keywords, and searched for them in the closed-caption of the stories broadcasted in the beginning and the end of his period in office.

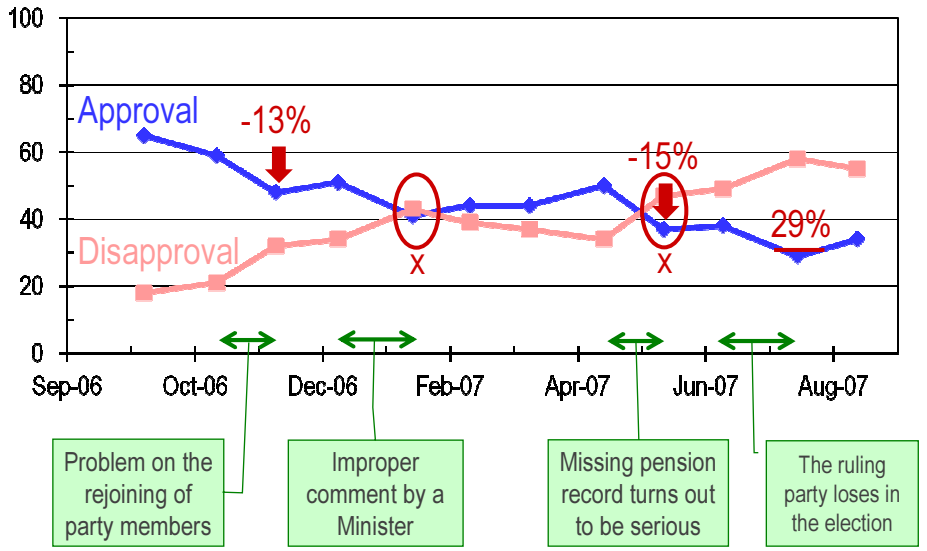

Fig. 2. An example of cabinet approval/disapproval rates and the detection of event periods in the case of ex-Prime Minister Shinzo ABE. When a dynamic behavior is observed in the graph, we consider that an important event occurred during the preceding period. The explanation in boxes are manually annotated to show the actual events that occurred. As a matter of fact, these events were mentioned in the corresponding Wikipedia article as causes of his resignation.

Topic-Based Event Detection. Since there are various events that could not be handled by the template-based approach, we decided to refer to the behavior of demoscopic polls. We referred to demoscopic polls provided monthly by NHK Broadcasting Culture Research Institute[4].

As shown in Fig. 2, we set the following conditions as drastic poll behaviors:

- Drastic increase / decrease in the approval rate (⇑/⇓).
- Reversal of approval and disapproval rates (X).
- Extremely high / low approval rate (—).

When either of the above conditions is observed, we considered that a major event occurred in the period of the current and the previous polls.

Next, news stories that describe the event during the period selected above are detected. In order to do so, we considered that the story should be either the beginning or the end of a news topic, or a heavily discussed story. We decided to measure such features from the topic thread structures proposed in [2]; A story is considered as a candidate that describes the major event if:

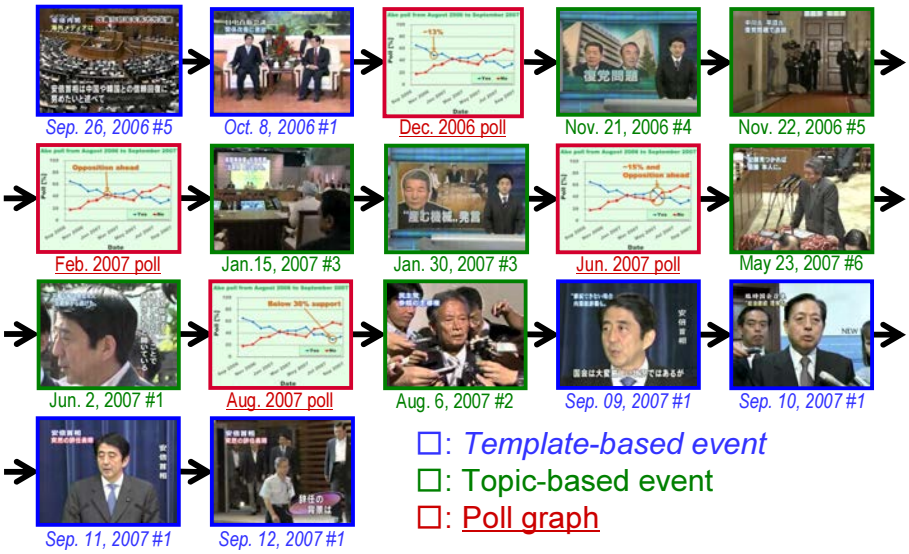

Fig. 3. Example of a generated summarized video in the case of ex-Prime Minister Shinzo ABE. Poll graphs are inserted when a drastic behavior is observed in the graph, followed by video clips that is supposed to contain descriptions on explanatory events (topic-based events). In the inauguration / resignation period, video clips are selected according to templates (template-based events).

– it is the beginning or the end of a topic thread structure, or
– stories are dense within a short period of time before and after it.

Finally, we apply sentiment analysis to the detected candidate stories referring to the dictionary created by Takamura et al. [5]. According to the user's preference, the candidates during a period are ranked in optimistic or pessimistic order, and selected from the most extreme ones.

Editing. In order to produce a summarized video with a specific length, we need to select a certain number of video clips with a certain length. We cropped a certain length of video segment starting from the utterance of the Prime Minister's name.

For the final output, poll graphs are inserted when a drastic behavior is observed in the graph, followed by video clips that is supposed to contain descriptions on explanatory events (topic-based events). In the inauguration / resignation period, video clips are selected according to templates (template-based events).

Fig. 3 shows an example of a summarized video produced by the proposed method.

3 Application to Sports Contents

For sports contents, we will introduce a work on video recompilation that makes use of the frequency and the estimated user attributes of twitter tweets as a "scenario".

3.1 Biased Sports Video Summarization According to Twitter Tweets

- Video: Sports show
- Scenario: Twitter tweets (frequency and estimated user attributes)

Recently, micro-blogging services such as twitter has become very popular. Especially in events such as sports games, it has become common for thousands of people to tweet while watching the game, sharing the experience in real time. Following this trend, we proposed a framework that analyzes tweets concerning a sports game (i.e. those with hashtags related to the game) to produce a summarized video biased towards supporters of each team, and developed an interface to demonstrate the results called "twiSpo" (Fig. 4). Details of the method could be found in [6].

Estimation of User Attributes. First, the attribute of each user who tweeted with a hashtag related to the game is estimated. Since it is difficult to analyze the attribute (i.e. which team the user supports) of each user from a single tweet, it is analyzed from a sequence of the user's tweets.

The classification was done by training keywords characteristic for tweets by users supporting each team. A dictionary for characteristic words was constructed by a method called SO-PMI (Semantic Orientation using Pointwise Mutual Information) [7].

Detection of Biased High-Light Scenes. Next, in order to detect high-light scenes biased towards each team, frequencies of tweets by supporters of each team are separately counted. Fig. 5 shows an example in the case of a baseball game. Based on thresholding to the frequency, summarized videos biased towards each team are produced.

4 Application to Cooking Contents

For cooking contents, we will introduce a work on video recompilation that makes use of text-based cooking recipes as a "scenario".

Fig. 4. The "twiSpo" interface. From the console on the top-right of the interface, users can select the team which he/she supports, and also the number and the duration of high-light scenes to be included in the summarized video. At the bottom, tweets are also shown along the summarized video played on the top-left.

4.1 Description of Text Recipes with Videos

- Video: Cook show
- Scenario: Cooking recipe

Recently, cooking recipe sites that allow posting from general users have become popular, and hundreds of recipes are posted to such sites on a daily basis. Although it is easy to post in text to such sites, it is still infrequent to post images and moreover videos that describe the cooking procedures due to the complexity of the editing.

Since cooking procedures are sometimes difficult to understand without visual description, we proposed a framework that automatically links corresponding video clips in a database to cooking operations in an arbitrary text recipe, and developed an interface called "Video CooKing" as shown in Fig. 6 to demonstrate the results. Details of the method could be found in [8,9].

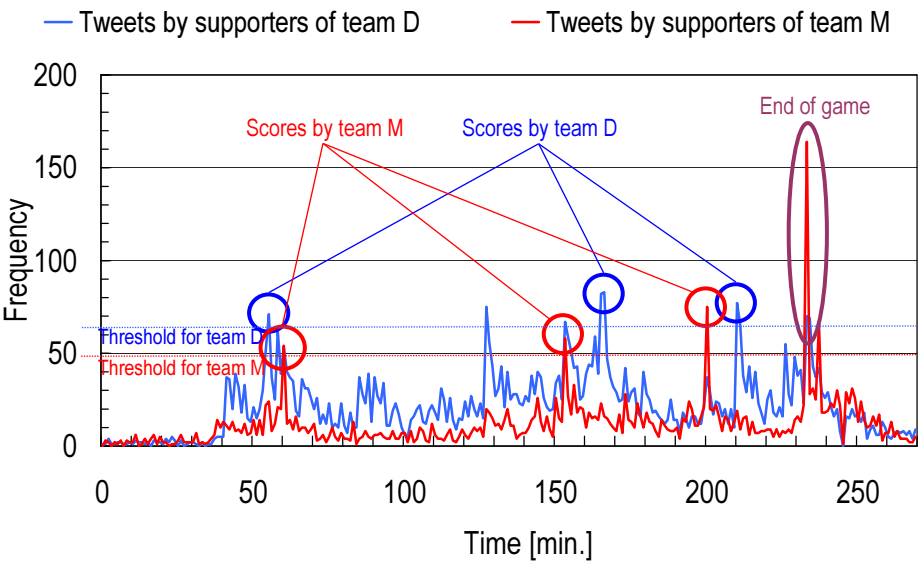

Fig. 5. Example of biased high-light scene detection. The two line graphs represent frequencies of tweets by supporters of each team. According to thresholds, high-light scenes are detected for each team. See the different high-light scenes detected.

Creation of a Video Database on Cooking Operations. In order to realize the framework, the most important part is the creation of a database that consists of video clips describing cooking operations. Since cooking operations may be different according to the ingredient to be cooked, the video clips should be annotated with a paired tag: (ingredient, operation).

Such a database could be created manually if possible, but due to the explosive number of combinations of ingredients and operations, we proposed a method to automatically create the database from cook shows broadcasted on TV. Most cook shows broadcasted in Japan come with closed-caption (audio transcript), so we analyzed the modification structure concerning imgredients that appear in it, in order to obtain (ingredient, operation) pairs that could be candidates for annotations.

However, the appearance in closed-caption does not always guarantee the existence of the actual cooking operation in the video. In addition, the video usually contains meaningless motions before or after the corresponding cooking operation. In order to correctly extract a video clip that describes a cooking operation corresponding to the (ingredient, operation) pair that appeared in the closed-caption around the same timing, motion features are analyzed.

As shown in Fig. 7, first, the motion in a video clip is classified into “repetitive”, “static”, or “others” by the trajectory of motion in the feature space. Next, repetitive motions are further classified in two by the distribution of repetitious motions in the frame. The motion class of the video clip is then matched with the

Fig. 6. The "Video CooKing" interface. The left side shows the original text recipe with links under cooking operations added by the proposed method. The right side shows video clips retrieved from the database which corresponds to the cooking operation specified in the text. There may be multiple corresponding video clips, so the user can learn different ways to perform the operation.

annotation candidates. In the end, if the classification matches, the annotation candidate is selected as the annotation for the video clip.

Linking Video Clips to a Text Recipe. When a text recipe is given, the modification structure concerning ingredients that appear in it is analyzed, in order to obtain (ingredient, operation) pairs. Next, the (ingredient, operation) pairs are sent to the video database as a query, and if available, corresponding video clips are linked from the text recipe. In case if there is no video clip corresponding to a certain (ingredient, operation) pair in the database, a partial match with only the operation is allowed.

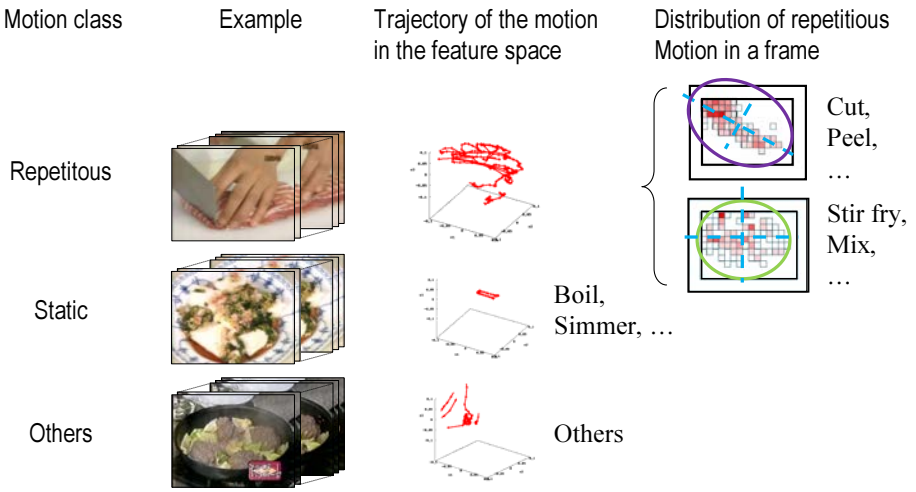

Fig. 7. Classification of cooking operations based on motion features. A video clip is classified first according to the trajectory of motion in a feature space, and then repetitive motions are further classified according to the distribution of motion in the frame.

5 Conclusion

In this paper, we introduced our works in news, sports, and cooking domains, that makes use of Wikipedia articles, demoscopic polls, twitter tweets, and cooking recipes in order to recompile video clips from corresponding TV shows.

In order to create contents bearable for practical use, besides improving the performance of individual techniques, we will need to collect and incorporate materials with a higher variety. We are considering to do so by establishing a framework that also makes use of videos available on the Web.

Acknowledgement. Parts of the works introduced in this paper were supported by Grants-in-aid for Scientific Research (B), for Scientific Research on Priority Areas (Infoplosion), and for Young Researchers (B), together with JSPS Excellent Young Researcher Overseas Visit Program, and joint research projects with National Institute of Informatics, Japan, and University of Amsterdam, the Netherlands. We would also like to thank the faculty and students involved in each work.

References

1. Okuoka, T., Takahashi, T., Deguchi, D., Ide, I., Murase, H.: Labeling news topic threads with Wikipedia entries. In: Proc. 11th IEEE Int. Symposium on Multimedia, pp. 501–504 (2009)

2. Ide, I., Kinoshita, T., Takahashi, T., Mo, H., Katayama, N., Satoh, S., Murase, H.: Efficient tracking of news topics based on chronological semantic structures in a large-scale news video archive. IEICE Trans. Information and Systems E95-D(5), 1288–1300 (2012)
3. Nack, F., Ide, I.: Why did the Prime Minister resign? —Generation of event explanation from large news repositories—. In: Proc. 19th ACM Int. Multimedia Conf., pp. 313–322 (2011)
4. NHK Broadcasting Culture Research Institute: The NHK Monthly Report on Broadcast Research. NHK Publishing, Inc., ISSN: 0288-0008
5. Takamura, H., Inui, T., Okumura, M.: Extracting semantic orientations of words using spin model. In: Proc. 43rd Annual Meeting of the Association for Computational Linguistics, pp. 133–140 (2005)
6. Kobayashi, T., Noda, M., Deguchi, D., Takahashi, T., Ide, I., Murase, H.: Summarizing sports video by on-the-spot comments on twitter (in Japanese). IEICE Technical Report, MVE2010-162 (2011)
7. Turney, P.D.: Thumbs up? Thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proc. 40th Annual Meeting of the Association for Computational Linguistics, pp. 417–424 (2002)
8. Doman, K., Kuai, C.Y., Takahashi, T., Ide, I., Murase, H.: Video CooKing: Towards the Synthesis of Multimedia Cooking Recipes. In: Lee, K.-T., Tsai, W.-H., Liao, H.-Y.M., Chen, T., Hsieh, J.-W., Tseng, C.-C. (eds.) MMM 2011 Part II. LNCS, vol. 6524, pp. 135–145. Springer, Heidelberg (2011)
9. Doman, K., Kuai, C.-Y., Takahashi, T., Ide, I., Murase, H.: Smart Video CooKing: A multimedia cooking recipe browsing application on portable devices. In: Proc. 20th ACM Int. Multimedia Conf. (to appear, 2012)