trackThem: Exploring a Large-Scale News Video Archive by Tracking Human Relations

Ichiro IDE[1,2], Tomoyoshi KINOSHITA[3], Hiroshi MO[2], Norio KATAYAMA[2], and Shin'ichi SATOH[2]

[1] Nagoya University, Graduate School of Information Science
Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan,
ide@is.nagoya-u.ac.jp
[2] National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan,
{ide|mo|katayama|satoh}@nii.ac.jp
[3] NetCOMPASS, Ltd.
#207, 5-17-8 Minami-Senju, Arakawa-ku, Tokyo, 116-0003, Japan,
kino@netcompass.co.jp

Abstract. We propose a novel retrieval method for a very large-scale news video archive based on human relations extracted from the archive itself. This paper presents the idea and the implementation of the method, and also introduces the *trackThem* interface that enables the retrieval and at the same time track down the relations. Although detailed evaluations are yet to be done, we have found interesting relations through the exploration of the archive by making use of the proposed interface.

1 Introduction

There have been many works aiming to retrieve news video contents. However, not much has been reported on what can be done and what can be acquired from a very large news video archive. In other words, most of the works aimed at analyzing and/or retrieving the contents in an archive either from individual units (shots, stories, and so on), or simply within each independent program.

Recent technology has enabled us to archive video data in large quantities. We have built an automatic broadcast video archiving system [Katayama *et al.* 2004], which has up to now, archived approximately 700 hours of daily Japanese news video spanning over the past four years. Even without such a system, large scale news video archives have become available, for example by participating to the TREC-Video workshop [NIST]. This has encouraged some groups to start exploring a news video archive according to its contents as a whole, in the aspect of 'Question and Answering' [Yang *et al.* 2003], 'topic threading' [Ide *et al.* 2004,Duygulu *et al.* 2004], and so on.

In this paper, we propose a topic-based news video browsing interface which provides access to an archive based on human relations extracted from the archive itself. This approach is proposed since human activity is the core contents in news videos, and that interactions between the humans cause such activities.

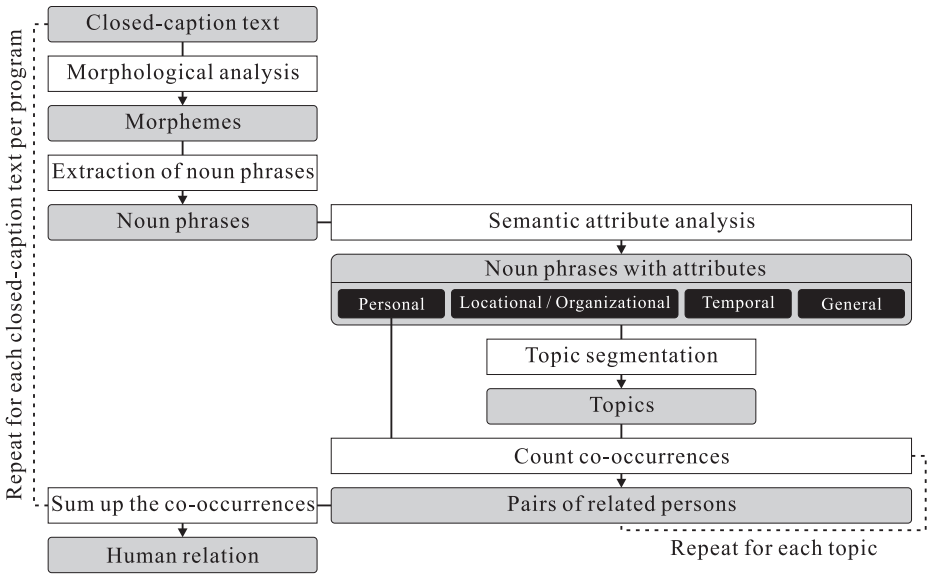

Fig. 1. Overall flow of the human relation extraction process.

Therefore, we consider that extracting human relations from a news video archive is an essential foundation for the understanding of its contents.

2 Extracting Human Relations from a News Video Archive

2.1 Overview

Extracting human relations from a large quantity of data has been a hot topic in the data mining, the semantic web, and the social network analysis fields [Kautz *et al.* 1997]. Works in these fields have focused mainly on e-mail correspondence [Golbeck *et al.* 2004], web links [Matsuo *et al.* 2005], references in academic papers [Yoshikane *et al.* 2004], and so on. However extraction of social networks from multimedia data has not been sought.

We are working on human relation extraction from a news video archive, in order to realize a retrieval and browsing interface based on the relations. As this enables a user to track down news stories according to human relations, the interface is named *trackThem*. It is generally considered that the so-called '5W1H' (*i.e.* When, Where, Who, What, Why, and How) are essential components of a news text in journalism, where '4W' (*i.e.* When, Where, Who, and What) are especially important entities to describe a story. Most previous works in news video retrieval field retrieve similar stories based on what has happened

('What'), some new works consider the time order ('When') as an important factor [Ide *et al.* 2004,Duygulu *et al.* 2004], and some other work provides an ability to retrieve/browse by geographical location ('Where') [Christel *et al.* 2000]. According to this classification, our proposal can be considered as a method based on 'Who', which has not been sought deeply in previous works.

In this Section, we will briefly present the process of the human relation extraction. The overall flow of the process is shown in Fig. 1.

2.2 Detecting Noun Phrases with a Human Attribute

Since news videos not only talk about famous people, but they also refer to nameless people, we would like to detect noun phrases that indicate persons (hereinafter personal nouns), including but not limited to names with proper nouns, as in named entity analysis.

Basically in Japanese language, the suffix determines the attribute of a noun phrase. Based on this nature, we have collected nouns that may represent humans either individually or as a suffix of a noun phrase.

As a pre-process, each sentence of a closed-caption text (transcript of the audio speech provided from the broadcaster) is analyzed by a Japanese morphological analyzer. [4] Next, noun phrases are extracted according to the morphemes, followed by semantic attribute analysis based on the collected nouns.

Details on the composing of the dictionary and the method can be found in [Ide *et al.* 1999]. According to evaluations applied to 2,549 super-imposed captions that appeared in news videos, a precision of 72.47% and a recall of 82.13% were achieved by this method.

2.3 Topic Segmentation

Topic segmentation is a major research topic in the text retrieval and the natural language processing fields. The aim of our work is not to compete with the existing works, so we applied a relatively simple method for the segmentation.

The following segmentation process is applied to each sentence of a closed-caption text:

1. Create keyword vectors for each sentence. Keyword vectors for four semantic attributes; general, personal, locational/organizational, and temporal, are formed by noun phrases that were extracted in Sect. 2.1. The latter two are analyzed in the same way with the personal noun by applying a different suffix dictionary to it, and all the others are classified as general nouns.
2. For each sentence boundary, concatenate w adjacent vectors on both sides of the boundary. Measure the similarity of the two concatenated vectors by calculating the cosine of the angle between them. Choose the maximum similarity among all the window sizes: w. The maximum w was set to 10 in the following experiment.

[4] JUMAN 3.61 distributed from Kyoto University was used as the analyzer.

3. Combine the similarities in each semantic attribute and detect a topic boundary when it does not exceed a threshold. According to a training with 384 manually given topic boundaries, we have obtained an optimal weight of 0.23 for general, 0.21 for personal, 0.48 for locational/organizational, and 0.08 for temporal nouns, and a threshold of 0.17.
4. Concatenate over-segmented topics by measuring the similarity of the keyword vectors between adjacent topics.

Details of this method can be found in [Ide *et al.* 2004]. According to previous evaluations applied to 384 manually annotated topic boundaries as a ground-truth, a precision of 90.5% and a recall of 95.4% were achieved by this method when we allowed mis-judgments at a maximum of ± 1 sentence.

2.4 Extracting Human Relations

We consider that persons that co-occur within a topic have some kind of relation. The relation $R(p_i, p_j)$ between two persons p_i and p_j is defined as follows:

$$R(p_i, p_j) = \sum\nolimits_t f(p_i, t) f(p_j, t) \ , \tag{1}$$

where t represents a topic, and $f(p, t)$ the frequency of personal noun p in topic t.

It might be effective to put a higher weight when two persons co-occur within a sentence, and gradually release the weights according to their distance within a topic. However this is left for future investigation. Grouping of the same person with different representation is another issue needed to be solved. However, the solution to this issue may not be trivial since some representations may be ambiguous and change along time.

3 Exploring the Archive by Tracking Human Relations

Currently, the process in Sect. 2 is running fully automatically every night after the broadcast of the program. As of June 5, 2005, we have 17,468 topics ranging over 1,454 days. Within these topics, there were 150,238 noun phrases with a human attribute, of which 15,686 were different. There were 307,951 edges, or human relations, between the 15,686 noun phrases.

Fig. 2 shows a part of the human relation graph structure. The graph shows strong bilateral edges (relations) between the nodes (persons) and also strong unilateral nodes that link only the nodes that appeared in the bilateral structure. This structure was extracted automatically by referring to the top ten edges from each node, except for the merger of different representations of the same person.

In order to provide the extracted relations for video retrieval, we developed an interface that enables a user to retrieve and view news topics related to a selected pair of persons, while at the same time, track down the human relations. A snapshot of the interface, namely the *trackThem* interface, is shown in Fig. 3. We believe that this should provide a better understanding of news topics from the view point of human relations.

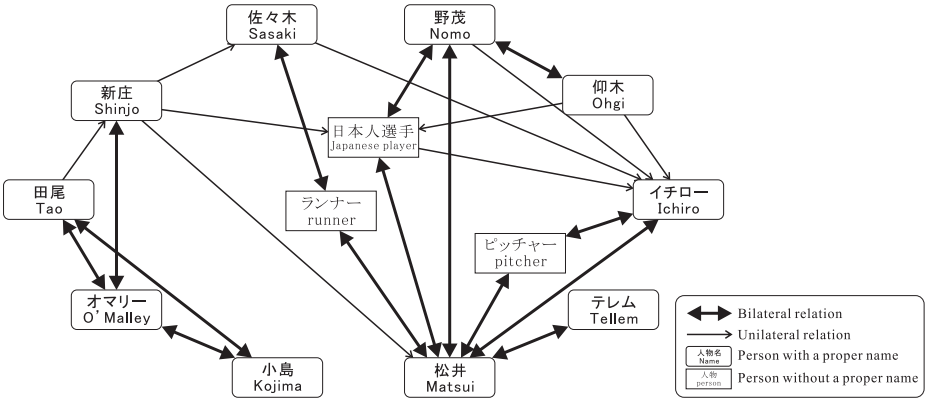

Fig. 2. A part of the human relation graph structure; the part mostly shows relations among Japanese baseball players playing in the US.

4 Conclusion

In this paper, we have proposed a novel approach for news video retrieval; retrieval and browsing according to human relations within a very large-scale archive. Future work includes the evaluation on the extracted relations and the efficacy of the interface. We are also considering to make use of co-occurrences of faces in the image, in order to extract relations that do not appear in the text.

Acknowledgements

This work was partly supported by a Joint Research program "Research on creating a large-scale video corpus" funded by the National Institute of Informatics.

References

[Christel *et al.* 2000] Christel, M.G., Olligschlaeger, M., and Huang, C.: Interactive maps for a digital video library. IEEE Multimedia **7(1)** (2000) 60–67

[Duygulu *et al.* 2004] Duygulu, P., Pan, J.-Y., and Forsyth, D.A.: Towards auto-documentary: Tracking the evolution of news stories. Proc. Twelfth ACM Intl. Conf. on Multimedia (2004) 820–827

[Golbeck *et al.* 2004] Golbeck, J. and Hendler, J.: Reputation network analysis for email filtering. Proc. First Intl. Conf. on Email and Anti-Spam (2004)

[Ide *et al.* 1999] Ide, I., Hamada, R., Sakai, S., and Tanaka, H.: Semantic analysis of television news captions referring to suffixes. Proc. Fourth Intl. Workshop on Information Retrieval with Asian Languages (1999) 37–42

[Ide *et al.* 2004] Ide, I., Mo, H., Katayama, N., and Satoh, S.: Topic threading for structuring a large-scale news video archive. Proc. Third Intl. Conf. on Image and Video Retrieval, Lecture Notes in Computer Science **3115** (2004) 123–131

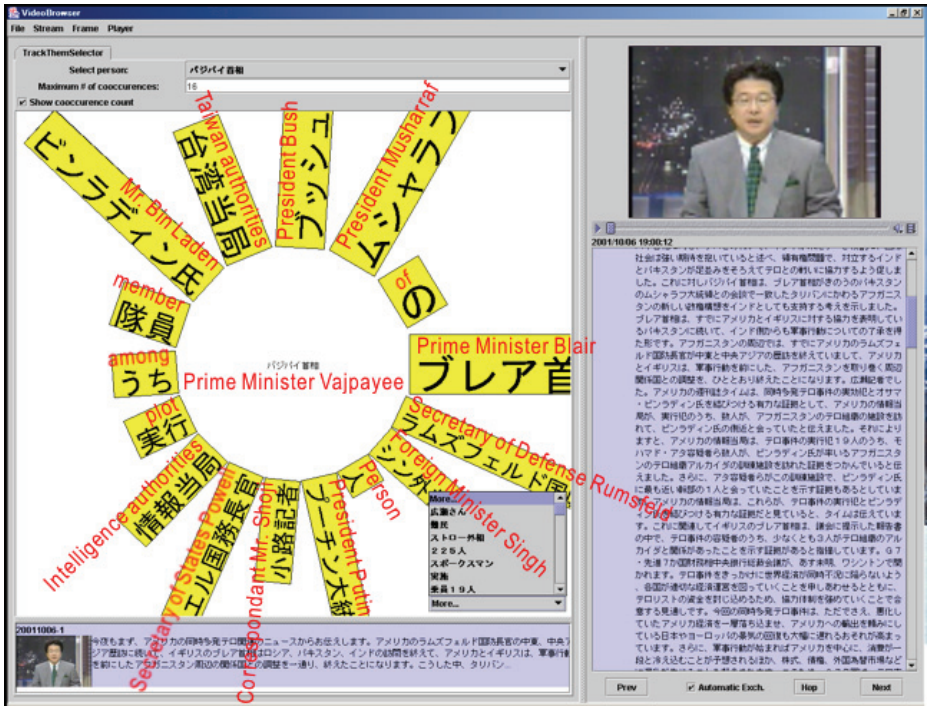

Fig. 3. The *trackThem* interface. The person in focus is encircled by related persons aligned in the order of their degrees of relation in proportional sizes. The person in focus may initially be selected from a list on the top. A single click on a related person enlists the topics that they co-occured at the bottom, while a double click switches the person in focus. A double click on a topic sets it on a video viewer on the right.

[Katayama *et al.* 2004] Katayama, N., Mo, H., Ide, I., and Satoh, S.: Mining large-scale broadcast video archives towards inter-video structuring. Proc. Fifth Pacific Rim Conf. on Multimedia, Lecture Notes in Computer Science **3332** (2004) 489–496

[Kautz *et al.* 1997] Kautz, H., Selman, B., and Shah, M.: The hidden web. AI Magazine **18(2)** (1997) 27–36

[Matsuo *et al.* 2005] Matsuo, Y., Mori, J., Asada, Y., Hasida, K., and Ishizuka, M.: Mining large-scale social network of researcher from the web. Proc. Fifteenth Intl. Sunbelt Social Network Conf., (2005)

[NIST] National Institute of Standards and Technology: TREC Video Retrieval Evaluation. `http://www-nlpir.nist.gov/projects/trecvid/`

[Yang *et al.* 2003] Yang, H., Chaisorn, L., Zhao, Y., Neo, S.-Y., and Chua, T.-S.: VideoQA: Question and answering on news video. Proc. Eleventh ACM Intl. Conf. on Multimedia (2003) 632–641

[Yoshikane *et al.* 2004] Yoshikane, F. and Kageura, K.: Comparative analysis of coauthorship networks of different domains: The growth and change of networks. Scientometrics **60(3)**, Akadémiai Kiadó, Budapest / Kluwer Academic Publishers, Dordrecht (2004) 435–446