



Interactive Story Maker: Tagged Video Retrieval System for Video Re-creation Service

Chang-Uk Kwak

Smart Media Research Group, Electronic
Telecommunications Research Institute
cukwak@etri.re.kr

Sun-Joong Kim

Smart Media Research Group, Electronic
Telecommunications Research Institute
kimsj@etri.re.kr

Min-Ho Han

Smart Media Research Group, Electronic
Telecommunications Research Institute
mhhan@etri.re.kr

Gyeong-June Hahm*

Smart Media Research Group, Electronic
Telecommunications Research Institute
hahm@etri.re.kr

ABSTRACT

Users who want to reuse existing videos and re-create new videos need a video retrieval system that searches relevant video clips from a vast amount of video clips. In addition, a re-creation tool is needed that allows users to arrange videos according to the user defined story line using the retrieved videos. Current video retrieval services are based on simple tags, they have potential limitations on searching relevant videos for natural language queries. That is, if you enter a scenario that describes a scene as a query, there is a limitation in getting the appropriate retrieval results. To address these problems, this paper introduces a system that performs query preprocessing and expansions and provides users with retrieval results for a natural language query. In addition, we introduce a web based re-creation tool that can construct story line using retrieved videos and play a re-created video.

CCS CONCEPTS

• Information systems → Video search;

KEYWORDS

video retrieval, video re-creation, query processing

ACM Reference Format:

Chang-Uk Kwak, Min-Ho Han, Sun-Joong Kim, and Gyeong-June Hahm. 2018. Interactive Story Maker: Tagged Video Retrieval System for Video Re-creation Service. In *2018 ACM Multimedia Conference (MM '18), October 22–26, 2018, Seoul, Republic of Korea*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3240508.3241401>

*Corresponding author

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MM '18, October 22–26, 2018, Seoul, Republic of Korea

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5665-7/18/10.

<https://doi.org/10.1145/3240508.3241401>

1 INTRODUCTION

As the video sharing platform expands, it is becoming more and more popular with contents re-creation by existing videos editing and arrangement. In order to find a desired part from a large amount of video clips, a video retrieval system is necessary to find a video corresponding to a natural language query such as 'a man and a woman are walking with each other'. Currently, the video retrieval systems provided by YouTube or portal sites are hard to find relevant videos because they are operated on naïve metadata tagging such as movie title with two or three keywords. That is, there is a limitation in retrieving a video including actor's action descriptions or dialogues corresponding to a natural language query. In this paper, we propose a system in which natural language user's query is preprocessed and expanded to provide relevant search results. In order to further enhance the usability of the video retrieval system, we introduce a tool that can arrange the order of the retrieved clips so that the user can easily re-create the story line on the web. This paper is composed as follows. Section 2 introduces the structure and implementation results of the video retrieval system. Section 3 explains the process of re-creating my own video easily and simply by constructing a story line using the retrieved videos.

2 TAGGED VIDEO RETRIEVAL SYSTEM

The proposed system constructs retrieval target videos by splitting original videos and entering their metadata to be retrieved. The system provides retrieval results by processing the user's natural language query. Figure 1 illustrates the overall structure and usage scenarios for the system. In this study, a retrieval system is developed for searching Korean movies. Since the run time of each movie is quite long, we divide a movie into scene units by using the automatic splitter module [5] which automatically recognizes scene boundaries using deep learning algorithm. Usually, a movie is divided into two to three minutes long scenes, and an average of 40 clips are generated from a movie. The actual script of the movie is used for generating metadata for each segmented movies. Scripts of target movies are analyzed to extract action and dialogue descriptions from scripts [3]. For segmented videos, metadata is generated using a corresponding part of the script. In the metadata, a portion of actual script text

corresponding to scene boundaries, as well as the name of character, the title of the movie and the location-landmark information appearing in the video are gathered. The pre-processing and the indexing for the constructed metadata are performed using Korean morpheme analyser and Lucene¹ open source search engine, respectively. In order to express user's search intention more precisely, we applied the results of previous studies [1] and [2] on query preprocessing and expansion. In other words, major keywords are identified through the parsing and rank analysis of the query, and the system query is generated by adding the related terms using our domain ontology[4] and word embedding model. Then, ranking process is performed by relevance measuring on the retrieval list obtained by Boolean matching of the system query, and the retrieval result is finally provided to the user. Our search engine is implemented based on Lucene, but the query preprocessing, expansion and the ranking part are newly designed and implemented. We have developed a search engine in the form of a RESTful API using the Spring framework² for general purpose. The proposed system is working on about 14,000 segmented videos for 150 Korean films, and Figure 2 shows the retrieval results for the example query on the web UI. When the query 'She also bowed in surprise, following him, who is deeply bent down' is typed, relevant videos are retrieved on the right side of the web UI. To evaluate the search engine performance, the nDCG measure is used. For the 20 sentences of queries, six raters evaluate the relevance of retrieved videos manually, and the system records 80.6% which is an acceptable level of performance.

3 VIDEO RE-CREATION SERVICE

To improve the utilization of the proposed video retrieval system, we develop a web-based video re-creation tool. By using nodes and edges, a skeleton of a story is constructed. Relevant videos are retrieved through the search system for each node, then new contents having different story lines can be re-created by utilizing existing videos. Figure 3 shows a screen shot of the video re-creation tool. Story branching is also supported, so interactive media can be created. Since the recent editing tools for creating interactive media such as Rapt Media³ or Racontr⁴ provide a simple and naïve search function, our system can help users to be competitive in their media creation activities.

ACKNOWLEDGMENTS

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [18ZH1300, Development of programmable interactive media creation service platform based on open scenario]

¹<https://lucene.apache.org>

²<https://spring.io>

³raptmedia.com

⁴racontr.com

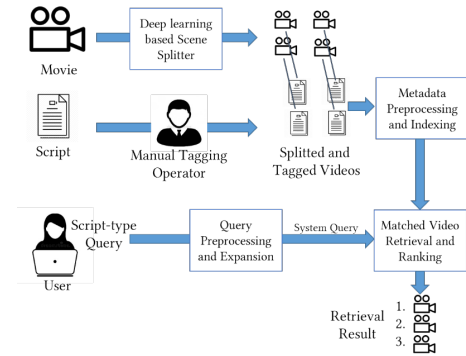


Figure 1: An overall process of the video retrieval system

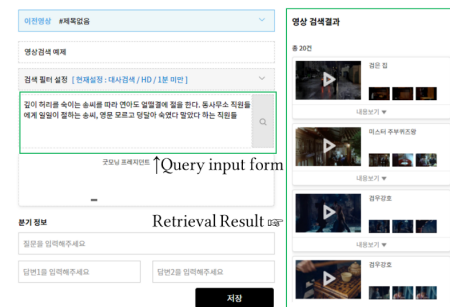


Figure 2: Video retrieval results for a natural language query

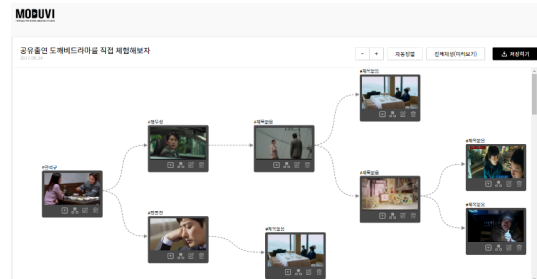


Figure 3: A screenshot of the web-based video re-creation tool

REFERENCES

- [1] G. Hahm, C. Kwak, and S. Kim. 2017. Tagged Video Retrieval System using Domain Ontology and Word Embedding. In *Proceedings of International Conference on ICT convergence*. 1101–1103.
- [2] C. Kwak, J. Son, Alex Lee, and M. Han. 2017. Scene Emotion Detection using closed caption based on Hierarchical Attention Network. In *Proceedings of International Conference on ICT convergence*. 1207–1209.
- [3] Alex Lee, J. Son, C. Kwak, and M. Han. 2017. Embedding Scene Segmented Video with Script. In *Proceedings of International Symposium on Advanced Intelligent Systems*. 1188–1191.
- [4] W. Park, M. Han, J-W. Son, and S-J. Kim. 2017. Design of Scene Knowledge based System based on Domain Ontology. In *Proceedings of the 19th International Conference on Advanced Communication Technology*. 560–562.
- [5] J. Son, C. Kwak, Alex Lee, and M. Han. 2017. Deep Ensemble Network for Movie Scene Boundary Detection. In *Proceedings of International Symposium on Advanced Intelligent Systems*. 1157–1161.