Recommendation Method to Make Combined Video from Video Segments

YunKyung Park, KyungDuk Moon
Electronics and Telecommunications Research Institute
Daejeon, South Korea
{parkyk, kdmoon}@etri.re.kr

Jungtaek Kim, Seungjin Choi
Pohang University of Science Technology
Pohang, South Korea
{jtkim,seungjin}@postech.ac.kr

***Abstract*—In this paper, we propose a recommendation scheme for media framework, which enables users to make their own videos by writing a story and reusing parts of accumulated videos. To reuse part of videos, we split a video into semantic segments based on an analysis of relations among objects in a video and store the segments with semantics in repository. To create a new video, the user sends queries based on his/her own story, then the framework recommends appropriate video segments for each query. To determine the rank of searched segments, the recommendation engine uses the degree of coincidence between the segment and the query. Also, it uses the degree of similarity between the searched segment and previously selected segment. By doing this, we can recommend a segment, which is consistent with user's intent and harmonized with the other parts of the new video.**

Keywords-component; Recommendation; Combined Video; Video Segment; Similarity; Conformity.

I. INTRODUCTION

Nowadays, it is common to take photos and shoot videos to record events in everyday life, and consequently visual data has been rapidly accumulating. Roughly 500 million photos are uploaded per day on social sites and 100 hours of videos per minute on YouTube. As visual data is accumulated, it has been reevaluated as new digital asset to trade and the most appropriate medium to represent one's thinking. This is because visual data can be intuitively understood, easily combined and transformed without any degradation of quality. To represent one's thinking, the parts of video should be selected and reconstructed freely by user. Also, new recommendation method is needed to minimize user effort.

Until now, recommendation algorithms are best known for their use on e-commerce Web sites [1], where they use input about a customer's interests to generate a list of recommended items. To recommend items, many applications use the items that customers purchase and explicitly rate to represent their interests, or use other attributes, including items viewed, demographic data, subject interests, and favorite artists [2]. Recommendation systems can be classified into two broad groups, content-based system and collaborative filtering system. Content-based systems examine properties of the items recommended. For instance, if a Netflix user has mainly watched cowboy movies, the system recommends a movie in cowboy genre. Collaborative filtering systems recommend items based on similarity measures between users and/or items. The items recommended to a user are those preferred by similar users [3]. In this paper, we propose a recommendation scheme for new media framework, which enables any user to make a new video through reusing parts of videos. The outline of this paper is as follows: in section II, we describe the proposed architecture in detail. In section III, we present the recommendation system and conclude in section IV.

II. MEDIA FRAMEWORK FOR COMBINED VIDEO

The proposed media framework enables the reuse of videos without user intervention. The framework consists of 4 subsystems: video analysis subsystem, search engine based on semantics, video segment recommendation subsystem and visualization subsystem which interfaces with the user. The video analysis subsystem analyzes the correlation of multi-modalities from an input video in order to precisely analyze the semantics of objects and splits a video into segments. The features of each modality are analyzed, after separating modalities from an input video. Noises and ambiguities of an object can cause an incorrect analysis. If each modality is only considered independently, it is difficult to overcome those factors, despite extensive efforts by researchers to address this. Video analysis subsystem then analyzes the correlations of modalities by fusion of modalities at the feature- and decision-level in order to overcome the effects of those factors [4][5]. Semantic search subsystem infers semantics of segment based on the objects and their trajectory in the segment, and stores the video segment with annotations. Moreover, semantic search subsystem searches all segments that matched user query.

The recommendation subsystem recommends a segment, which is the most appropriate segment to make a new video, among searched results. Many studies on recommendation, mainly confined to movies, consider not only information of persons, such as demographic data and behavior pattern, but also information on related content and persons, such as content with a similar subject or categories, and information on social relations [6][7]. Since the framework makes a new video using segments from different videos, the similarity among segments and the development of the story should be considered for a better recommendation. To choose the most appropriate segment, it considers semantic accuracy and similarity of back and forth segments. The visualization subsystem provides the user interface for describing a story, and dynamically reconstructs a series of segments according to the story. The subsystem extracts keywords and segment-sequence from the user story. Next it sends keywords to the search engine, and receives metadata for the recommended segment. Using Uniform Resource Identifier (URI) of original video, start and end point of the segment in the metadata, it makes video plot. The player plays the video according to the plot.

III. Recommendation System

We determine the proper segment through a 3-step process. First, when the subsystem receives searched segments from the search engine, the subsystem filters the segments based on some contexts in profile, such as guidance rate, using device, user defined attributes, etc. Second, it analyzes semantic correspondance between query and searched segments, and similarity between previously selected segment and searched segment. To reflect user preference, user ranks are predicted and used as weights. Third it ranks the segments depend on previous learning, how much the factors effect user selection.

Each query consists of <Subject, Predicate, Object> triples which describe the segment, Action State (AS) which describes movement direction of the objects and some user Designated Features (DF). To compute the semantic correspondance between the query and the searched segments, we created vector space for SPO, DF and AS. To compute degree of correspondance, we used euclidean distance in SPO, DF vector space and cosine similarity in AS vector space. Fig. 1 shows an example of semantic correspondance.

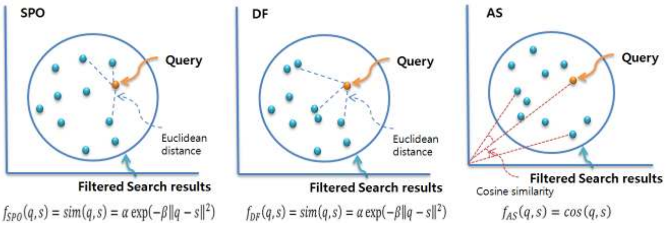

Figure 1. Example of a Semantic Correspondance.

The similarity between previously selected segment and searched segment is computed using euclidean distance in SPO space, Low-level Features (LF) space and cosine similarity in AS space. Fig. 2 shows an example of similarity between previously selected segment and searched segment.

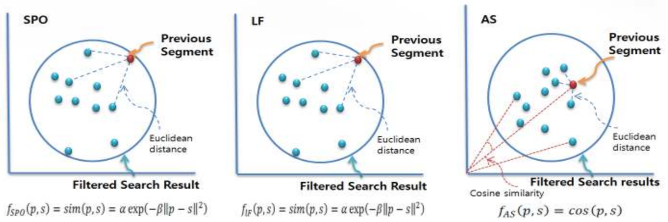

Figure 2. Example of a Similarity Between Back and Forth Segments.

In filtered search results, some segments have user rank and other segments have no rank. To predict the user rank for unranked segments, we propagate the known rank to unknown rank using a matrix, which is composed of rank propagation probability based on similarity. Fig. 3 shows an example of preference prediction using distance. In the figure, *lspo, las, ldf, llf* are predicted preference values for unlabeld segment in each space. They show how much the user is going to like the segment. The values are computed using label propagation algorithm. As a result, every segment has forur preference values.

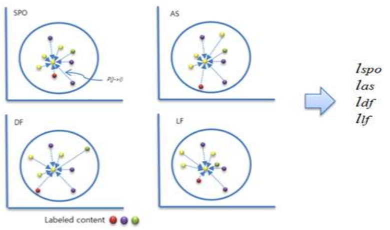

Figure 3. Example of Preference Prediction Using Distance

The final score of a segment is determined by weighted semantic correspondance and similarity. In equation (1), *wqi* and *wpi* are weights which are learned from user history. And *fspace(q,s), fspace(p,s)* are consistency and similarity measure in each space, respectively.

$$S = w_{q1} l_{spo} f_{SPO}(q,s) + w_{q2} l_{df} f_{df}(q,s) + w_{q3} l_{as} f_{as}(q,s) + w_{p1} l_{spo} f_{spo}(p,s) + w_{p2} l_{lf} f_{lf}(p,s) + w_{p3} l_{as} f_{as}(p,s) \quad (1)$$

IV. Conclusion and Future Works

Here, we propose a recommendation scheme for media framework, which enables users to create their own video by analyzing semantics of video and combining segments from different videos. In this recommendation system, the key differences are that it uses similarity between previously selected segment and recommending segment, and consistency between query and recommending segment. It can diminish awkwardness that occurs when the framework combines segments. To customize the recommendation, more research is needed on implicit user behavior. To assess the developed subsystem, more research on the evaluation method is also needed.

Acknowledgment

This work was supported by the ICT R&D program of MSIP/IITP. [R0126-15-1112, Development of Media Application Framework based on Multi-modality which enables Personal Media Reconstruction].

References

[1] J. B. Schafer, J. A. Konstan, and J. Reidl, “E-Commerce Recommendation Applications” Data Mining and Knowledge Discovery, Kluwer Academic, pp 115-153, 2001.

[2] G. Linden, B. Smith, and J. York, “Amazon.com recommendations item-to-item collaborative filtering”, IEEE Internet Computing, pp 76-80, January • February 2003.

[3] A. Rajaraman and J. D. Ullman, “Mining of Massive Datasets”, Cambridge University Press, 2011.

[4] R. Frcke, J. Thomsen, "LinkedTV Platform and Architecture", LinkedTV, 2012.

[5] P. K. Atrey, M.. A. Hossain, A. E. Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: a survey", Multimedia Systems, Volume 16, Issue 6, pp345-379, 2010.

[6] W. W. Cohen, "Collaborative Filtering: A Tutorial,"Carnegie Mellon University, https://www.cs.cmu.edu/~wcohen/collab-filtering-tutorial.ppt, [retrieved: Aug, 2016].

[7] D. Tsatsou et al., "Specification of user profiling and contextualization", LinkedTV, 2012.