Hyper Video Browser: Search and Hyperlinking in Broadcast Media

Maria Eskevich, Huynh Nguyen, Mathilde Sahuguet, Benoit Huet
EURECOM, Sophia Antipolis, France
huet@eurecom.fr

ABSTRACT

Massive amounts of digital media is being produced and consumed daily on the Internet. Efficient access to relevant information is of key importance in contemporary society. The Hyper Video Browser provides multiple navigation means within the content of a media repository. Our system utilizes the state of the art multimodal content analysis and indexing techniques, at multiple temporal granularity, in order to satisfy the user need by suggesting relevant material. We integrate two intuitive interfaces: for search and browsing through the video archive, and for further hyperlinking to the related content while enjoying some video content. The novelty of this work includes a multi-faceted search and browsing interface for navigating in video collections and the dynamic suggestion of hyperlinks related to a media fragment content, rather than the entire video, being viewed. The approach was evaluated on the MediaEval Search and Hyperlinking task, demonstrating its effectiveness at locating accurately relevant content in a big media archive.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems; H.5.4 [**Information Interfaces and Presentation**]: HyperText/Hypermedia

Keywords

Multimedia Search, Media Fragment, Hyperlinking

1. INTRODUCTION

Video traffic steadily increases to the level of representing two thirds of Internet communications within the next two years[1]. The number of online media platforms featuring professional (e.g. broadcast) and user contributed content is rising, together with the amount of media content available for consumption. With the sheer volume of data available, finding relevant information becomes a challenge, as in the case of videos, the search is less straightforward due to the lack of interpretability of their content. New online tools providing flexible, on-demand and relevant media documents are urgently needed to help users navigate in the media maze that a large-scale collection represents. To this end, we propose a Hyper Video Browser allowing for fine-grained media browsing, searching and hyperlinking within video collections. Fine-grained access to media content is achieved by computing and indexing descriptors and metadata at both the whole document and the media fragment level[2]. The proposed system may also be seen as an integrated second screen demonstrator, allowing viewers to actively search for additional content, cf. Figure 2, and proposing/recommending content related to the media fragment the user is watching dynamically, cf. Figure 3. It differs from common video search engines (i.e. YouTube, Dailymotion) by operating at finer-temporal scale and by taking video content into account throughout the process. The dataset from the MediaEval 2014 Search and Hyperlinking task [2], containing 2323 broadcast programs from various genre and totaling 1697 hours of content has been processed by our system as presented in Section 2. These media files are indexed and processed as a collection of 76240 topically coherent media fragments in order to become entirely available for users to browse, search and navigate though hyperlinks based on their content as detailed in Section 2.3.

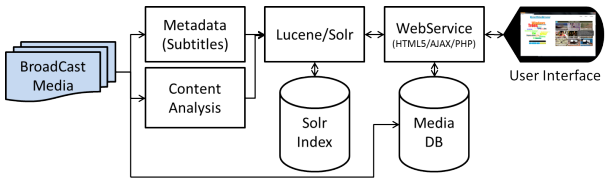

Figure 1: System Architecture Overview

2. SYSTEM ARCHITECTURE

The "Hyper Video Browser" architecture, as depicted in Figure 1, is composed of both offline and online processing components. Multimodal content analysis and indexing (using Lucene/Solr[3]) is performed offline, whenever a new video is added to the archive. A Web-service issues queries to the Solr index at run-time corresponding to the user activity.

2.1 Content Analysis

When a new multimedia document is ingested into the Hyper Video Browser, it is stored into the media database and processed at two levels: the entire document, where various metadata are available, i.e. title, cast, description, broadcast

[1]Cisco Visual Networking Index

[2]W3C Recommendation

[3]http://lucene.apache.org

MM'15, October 26–30, 2015, Brisbane, Australia.
ACM 978-1-4503-3459-4/15/10.
http://dx.doi.org/10.1145/2733373.2812618.

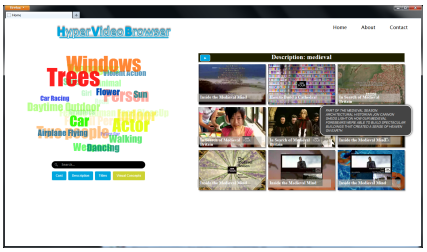

Figure 2: The Search/Browse Interface

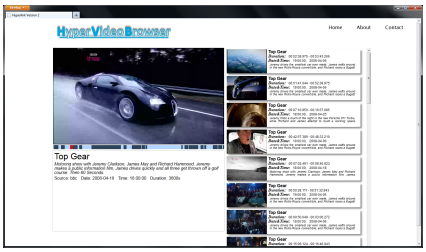

Figure 3: The Hyperlinking Interface

time, and the media fragment. The media fragments are defined using a complex segmentation procedure. Firstly, each video is segmented into shots. Then subtitles corresponding to each shot are extracted. Visual concept detection and optical character recognition are carried out on the key-frame associated to each shot [3]. Finally, the shots are grouped into larger segments (scenes or media fragments), based on visual similarity and temporal consistency [4].

2.2 Content Indexing & Searching

Both the metadata available with the broadcast and the extracted content are entered in a Lucene/Solr index at different granularities: video/scene/shot levels, allowing for precise jump-in point to be detected. Multimedia documents are represented by textual fields (for the textual metadata) and floating point fields (for the visual concept detector responses). When searching using text query, Solr's default TF-IDF-based text search is performed. When searching for visual concepts, the query is composed of range queries on the corresponding visual concepts. When searching for visual information using free text queries, the relevant visual concepts are identified based on their semantic WordNet distance with the query keywords. Multimodal queries are achieved by extending the textual query with range queries on the values of the selected visual concepts.

Hyperlinking is accomplished by automatically crafting a multimodal query from the currently played media fragment. The text query is compiled by extracting keywords from the subtitles aligned between the start and end time of the media fragment. Visual concepts scores are taken from the corresponding indexed data of the media fragment keyframes. If the media fragment contains more than one shot, the highest score over all shots for each concept is used.

2.3 Novel Media Browsing Interface

The interface has been designed with simplicity in mind and is composed of only two views: "Search and Browse" versus "Hyperlinking".

Figure 2 depicts the Hyper Video Browser home screen which is devised for browsing and searching for relevant media within the collection. It is composed of two main parts. The left hand side, provides rapid textual access to the media archive. It features a clickable animated tag-cloud, which can be selected to visualise the dataset content through various facets: Cast, Title, Description and Visual Concepts in our case. Other dimensions, such as Genre could easily be added along with one showing the user's favorite tags. Clicking on a tag, issues a textual search for the corresponding keyword/concept, which results in the presentation of a ranked list of media fragments. It is also possible to issue a free text query using the search field. Following the MediaEval Search and Hyperlinking scenario, the query may contain an additional field describing the visual content of the target video (e.g. 'visual:="visual description"'). The right hand side of the interface is contextualized so that a selection of 9 media fragments corresponding to a keyword randomly selected from the tag-cloud is displayed, providing a visual entry point to the collection, until the user issues a request. When a request is made, a ranked list of media fragment is shown (keyframe and metadata).

Upon selecting a video from the main screen, the user enters the second view panel: the hyperlinking player. Figure 3 shows a list of hyperlinks whose content relates to the content of the video fragment currently playing (on the left). When a new scene (media fragment) is reached, the system automatically constructs and issues a query to the Solr index for related content, and the hyperlinks are updated. Selecting a hyperlink results in its playing and an update in relevant hyperlinks suggestions on the right. A textual description of the program and the current fragment is dynamically displayed below the player. A bi-colored temporal stripe underneath the displayed video gives direct access to individual media fragments and provides insight upon the program structure. These functionalities empower the viewer to explore and reach media content more effectively.

3. CONCLUSION

The Hyper Video Browser endeavors to facilitate user access to multimedia content. The novel media browsing interface provides effective skimming using categorized tag-clouds, allows to perform textual and visual searches (using natural language queries) and offers hyperlinks to media containing related content dynamically while playing the selected video. The system performed among the best of MediaEval Search and Hyperlinking task, indicating the relevance and accuracy of the system [1].

4. ACKNOWLEDGEMENTS

This work has been partially supported by Bpifrance within the NexGen-TV Project, under grant number F1504054U.

5. REFERENCES

[1] E. Apostolidis, V. Mezaris, M. Sahuguet, B. Huet, B. Cervenková, D. Stein, S. Eickeler, J. L. Redondo Garcia, R. Troncy, and L. Pikora. Automatic fine-grained hyperlinking of videos within a closed collection using scene segmentation. In *ACMMM 2014, 22nd ACM International Conference on Multimedia*, Orlando, Florida, USA, 11 2014.

[2] M. Eskevich, R. Aly, D. N. Racca, R. Ordelman, S. Chen, and G. J. Jones. The Search and Hyperlinking Task at MediaEval 2014. In *Proceedings of MediaEval 2014 Workshop*, Barcelona, Catalunya, Spain, 2014.

[3] H. Le, Q. Bui, B. Huet, B. Cervenková, J. Bouchner, E. E. Apostolidis, F. Markatopoulou, A. Pournaras, V. Mezaris, D. Stein, S. Eickeler, and M. Stadtschnitzer. LinkedTV at MediaEval 2014 Search and Hyperlinking Task. In *Proceedings of MediaEval 2014 Workshop*, Barcelona, Catalunya, Spain, 2014.

[4] P. Sidiropoulos, V. Mezaris, I. Kompatsiaris, H. Meinedo, M. Bugalho, and I. Trancoso. Temporal video segmentation to scenes using high-level audiovisual features. *IEEE Trans. Circuits Syst. Video Techn.*, 21(8):1163–1177, 2011.