The VERGE System at the Video Browser Showdown 2018 - Final Notes

Anastasia Moumtzidou[1], Stelios Andreadis[1], Foteini Markatopoulou[1,2], Damianos Galanopoulos[1], Ilias Gialampoukidis[1], Stefanos Vrochidis[1], Vasileios Mezaris[1], Ioannis Kompatsiaris[1], Ioannis Patras[2]

[1] Information Technologies Institute/Centre for Research & Technology Hellas, Thessaloniki, Greece
{moumtzid|andreadisst|markatopoulou|dgalanop|heliasgj|stefanos|bmezaris|ikom}@iti.gr

[2] School of Electronic Engineering and Computer Science, QMUL, UK
i.patras@qmul.ac.uk

ABSTRACT

This paper provides further details of the VERGE system that was developed by team ITI_CERTH for the VBS2018 [2]. It includes an overview of the system's search features, the changes made after the publication of the VBS2018 paper [1], and a basic analysis of how the system has been used for VBS2018 from a user perspective.

KEYWORDS

video browser showdown, evaluation campaign, ad-hoc video search

1 SYSTEM OVERVIEW

VERGE system (Figure 1) has established a solid presence during the last years of the VBS [2]. The application aims at providing a user-friendly interface that incorporates various content retrieval features, in order to search efficiently through large collections of video frames. The latest version of VERGE (http://mklab-services.iti.gr/vbs2018), which participated in VBS2018 [1], includes:

- a visual similarity search feature for query-by-example;
- a high-level visual concept-based search;
- a free-text query conversion to high-level visual concepts;
- a color filtering of 24 dominant colors;
- a video clustering module;
- a text search module in video level exploiting metadata;
- a video similarity module combining concepts and metadata.

The search results are shown as image shots in a grid view, sorted by retrieval scores. The system provides an option to either retrieve fresh results or re-rank the already resulted shots. Also, a film-strip displays the complete set of video shots for a selected frame.

2 FINAL SYSTEM CHANGES

The system described in VBS2018 paper [1] has been further developed after paper submission, resulting to a few changes. First, the text search module was limited to using simple text from video metadata. In detail, after internal evaluations, it was decided that the ASR transcripts would not be considered due to noise and DBPedia concepts would be replaced by stems of the original text using Apache Lucene. Another change concerns the multimodal fusion module, where a simpler approach was followed that fuses data on video level and considers visual concepts and textual metadata. Each video is represented with the top-K detected concepts of its video keyframes, then Jaccard Similarity Coefficient is used to find similar videos for the two modalities, and, finally, the results are merged by ranking higher videos with a better score in both lists.

Figure 1: Screenshot of VERGE video retrieval engine.

3 SYSTEM USAGE AT VBS2018

Varied components were favored accordingly to different types of the contest's queries, i.e. visual KIS, textual KIS, and textual AVS. During textual tasks the users employed mostly concept search, since no or few clues about the visual representation of the scene were provided. The option of combining more than one concepts was proved to be of great value and the mapping of free text to existing concepts significantly enhanced the search. There were cases that re-ranking of results by color was also substantial to reach the final goal. Especially for the AVS queries, when users reached a relevant keyframe, they applied visual similarity in order to retrieve more results. Naturally, during visual tasks, color search and visual similarity played a leading role. All in all, video-based clustering and multimodal fusion on video level were not found very helpful.

ACKNOWLEDGMENTS

This work was supported by the EU's Horizon 2020 research and innovation programme under grant agreements H2020-645012 KRISTINA, H2020-779962 V4Design, H2020-732665 EMMA, and H2020-687786 InVID.

REFERENCES

[1] A. Moumtzidou, S. Andreadis, F. Markatopoulou, D. Galanopoulos, I. Gialampoukidis, S. Vrochidis, V. Mezaris, I. Kompatsiaris, and I. Patras. 2018. VERGE in VBS 2018. In *MultiMedia Modeling - 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part II.* 444–450. https://doi.org/10.1007/978-3-319-73600-6_48

[2] K. Schöffmann and W. Bailer. 2012. Video Browser Showdown. *SIGMultimedia Rec.* 4, 2 (July 2012), 1–2. https://doi.org/10.1145/2350204.2350205