Supporting Browsing of User Generated Video on a Tablet

Frank Hopfgartner, David Scott, Jinlin Guo, Yang Yang, Cathal Gurrin, Alan F. Smeaton
CLARITY: Centre for Sensor Web Technologies
Glasnevin, Dublin 9, Ireland
{fhopfgartner, dscott, jguo, yyang, cgurrin, asmeaton}@computing.dcu.ie

ABSTRACT

In this demo paper, we describe our user-generated video search system, compromising of an iPad interface communicating with a remote server. The goal of this system is to provide an easy access to video content lacking textual annotations by clustering key frames. Moreover, the graphical user interface allows users to filter video content based on various semantic concepts.

Categories and Subject Descriptors

H.5 [**Information Systems Applications**]: Information Interfaces and Presentation

General Terms

Design

Keywords

multimedia information retrieval, tablet, interface

1. INTRODUCTION

In recent years, one could observe the increasing success of so-called Web 2.0 applications, i.e. online platforms that live from user generated content. Nowadays, more and more people do not only actively consume content online, they have also started to create their own content, e.g. on online platforms such as Wikipedia, Facebook and YouTube. According to recent statistics[1], people now upload more video content every day than the three major US American networks created within the past 60 years. Various challenges arise from this rather un-coordinated development. First of all, we are facing large amounts of user generated video that, given different motivation, skills and experience of contributors, can not easily be categorized [1]. User-generated videos are very diverse, as can be seen in the different quality, length and topics of these videos. Moreover, metadata information such as speech transcripts or other textual clues are often not available.

An important challenge is to support users in accessing such large data collections containing unstructured and sparsely annotated video data [2]. In [4], we apply concept detectors to provide a basic semantic annotation of video content. Preliminary experiments indicate that concepts can successfully be used to identify simple events in video data. In this demo, we will indicate how such concepts can be used to filter video content, thus easing users' access to the data collection. In Section 2, we first introduce the data collection that we use to demonstrate our system. Then, we discuss how we process these video data in Section 3. In Section 4, we present the graphical user interface. Section 5 concludes this work.

2. DATA CORPUS

For this demo, we employ the NIST TRECVid 2011 Multimedia Event Detection (MED) corpus [5]. The MED task was to foster automatic complex event recognition in internet video. The internet video was collected by the Linguistic Data Consortium and consists of publicly available, user-generated content posted on various Internet video hosting sites. Given the high reputation of TRECVid in multimedia retrieval research, we consider this data corpus to be an appropriate snapshot of user-generated content.

3. DATA PRE-PROCESSING

We devised a video retrieval system which incorporated visual clustering techniques to present search results as a ranked list of video clusters. This allows the users to view visually similar content clustered together, and reduces the overhead of scrolling/browsing through the whole ranked list. The system accepts the selection of visual concepts as the query mechanism and operates on a tablet PC with the aim of providing the user with a simple interface to a complex back-end video search tool. Most of the search functionality and video content processing for indexing and presentation (outlined below) take place on a server. A detailed description of the system's backend is provided in [4].

Accurate key frame selection is especially influential given that our system is heavily focused on the video ranked result representation. We employ two types of key frame selection criteria, firstly the "most average keyframe" is chosen using the MPEG-7 descriptors, Edge histogram, Color Layout and Scalable Color, and secondly we employ a query-biased keyframe selection approach when the user has entered vi-

[1]`http://www.youtube.com/t/press_statistics`, accessed on 27 January 2012

ICMR'12, June 5–8, 2012, Hong Kong, China.

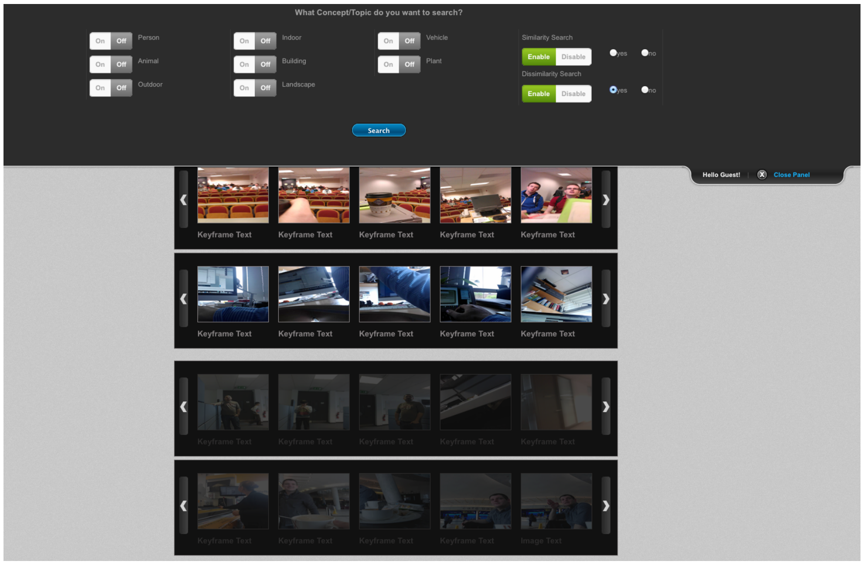

Figure 1: Screenshot of the Tablet interface

sual concepts to identify query-appropriate keyframes. For cases when a single visual concept is included, the top-ranking key frame (for that concept) is chosen; in the case where more than one concept is selected evidence from all concepts are fused to identify the top-ranked frame.

4. INTERFACE

Differing from most video retrieval interfaces ([3]), our chosen platform is a tablet PC, which can be either an iPad, WebOS or any Android tablet. Figure 1 shows a screenshot of the graphical user interface which has been developed in HTML 5, thereby allowing cross-platform deployment. The user is presented with an interface which has a title bar (top of the screen) containing a set of concepts that help to partition the collection. Users select concepts to build a visual query, this visual query returns a ranked list of shots which are displayed in descending order of relevance. The user can either select a key frame to determine if it is correct, select other concepts to refine the search or do a similarity search on the selected key frame to obtain items with similar visual, textual features or key frames from the same video segment. At the point when the user has found the required video segment, they will tag the video segment and move onto process another information need.

5. EVALUATION AND CONCLUSION

In a preliminary evaluation, we asked six users, four novice users and two expert users, to use our video retrieval system following a search scenario as outlined within TRECVid. The four novice users were students from a business school with English was their second language. The novice users had never seen either system before and had limited experience of using a tablet PC device.The two expert users were not directly involved in the system but had experience in video search. Overall user satisfaction was positive based on a post-experiment user evaluation.

In order to evaluate the clustering technique of our system, we asked our participants to use two different interfaces: A baseline system where each entry in the search results list represents a video, and the cluster-based system as described above. The clustering system outperformed the baseline with regard to Mean Elapsed Time with the novice users taking an average of 2.66 minutes per topic for the clustering system and experts taking 3.022 minutes for the experts clustering; the novices using the baseline system took an average of 3.324 minutes per topic. In our baseline experiments the novice users found a total of 12 topics out of the 25, our novices using clustering found one more topic at 13 and our expert users found one more again to give us a total of 14 out of 25 found. Summarizing, our initial evaluation reveals the effectiveness of our system.

Acknowledgements

This research was supported by the Norwegian Research Council (CRI number: 174867) and Science Foundation Ireland under Grant No. 07/CE/I1147.

6. REFERENCES

[1] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *IMC'07: 7th ACM SIGCOMM Conference on Internet Measurement*, pages 1–14, New York, NY, USA, 2007. ACM.

[2] F. Hopfgartner. *Understanding Video Retrieval*. VDM Verlag, 2007.

[3] K. Schoeffmann, F. Hopfgartner, O. Marques, L. Boeszoermenyi, and J. M. Jose. Video browsing interfaces and applications: a review. *SPIE Reviews*, 1(1):018004, 2010.

[4] D. Scott, J. Guo, C. Foley, F. Hopfgartner, C. Gurrin, and A. F. Smeaton. TRECVid 2011 Experiments at Dublin City University. In *TRECVid 2011*, 12 2011.

[5] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. In *MIR '06: 8th ACM Int. Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press.