Active Learning Approach to Interactive Spatio-temporal News Video Retrieval

Huan-Bo Luan[1], Shi-Yong Neo[2], Tat-Seng Chua[2],
Yan-Tao Zheng[2], Sheng Tang[1], Yong-Dong Zhang[1], Jin-Tao Li[1]
[1]Institute of Computing Technology, CAS [2]School of Computing, NUS
{hbluan, ts, zhyd, jtli}@ict.ac.cn {neoshiyo, chuats, yantaozheng}@comp.nus.edu.sg

ABSTRACT
Interactive news video retrieval requires the effective communication between the human searchers and the search engine to locate relevant video segments. We propose a spatio-temporal visual map (STVM) retrieval [1] system coupled with active learning to support user-centered interactive retrieval.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]: Retrieval Models, Search Process

General Terms
Design, Experimentation

Keywords
Interactive Retrieval, Active Learning

1. INTRODUCTION
In this paper, we present a framework based on our proposed STVM model (http://go2view.ict.ac.cn), which incorporates the sophisticated spatio-temporal visual relationship on top of the text-based retrieved results to enhance the overall retrieval performance. This type of "visual + text" retrieval has been leveraged in many content-based retrieval systems and found to be effective. An interactive user-interface is designed based on the above methodology to permit users to browse videos according to the spatio-temporal patterns in news video and provide feedback. To enhance the performance, we further integrate active learning into the framework. This active learning approach will leverage feedback from the user to re-assign weights to relevant features that are deemed to be important to user's query.

2. SEARCH BASELINE
We make use of the NUS's fully automated search system [2] submitted to TRECVID 2006 [3] as our search baseline. The system exploits: (a) the use of semantic visual concepts by performing query-analysis to relate user's query to available high-level features (HLFs); and (b) the integration of event structures present in news video for event-based retrieval. The proposed framework utilizes various multimodal features such as Video OCR, name entities present in automated speech recognition (ASR) text, and HLFs for news video retrieval.

3. SPATIO-TEMPORAL VISUAL MAP
The STVM uses the intuition that a relevant video clip S_c has several visually dissimilar keyframes and this information can be leveraged to re-rank other video clips S_r in the corpus that have keyframes which are similar to any keyframe of S_c. The scoring function will propagate the score of relevant clips S_c to other video clips S_r. The final score of S_r is obtained by combining the context similarity (between S_r and text query) and visual similarity (between keyframes of S_c and S_r).

4. ACTIVE LEARNING
Relevance feedback is a critical component of our system and it uses the relevant shots tagged by the users to enhance back-end re-ranking. We apply an active learning framework [3] based on support vector machines (SVM). The learning algorithm learns the supportive and non-supportive features using the set of shots tagged by the user. This is followed by a real-time re-calculation of the scores of shots that are not yet labeled. Finally, the system will present the most probable set of shots to the users for relevance feedback.

5. INTERACTIVE USER INTERFACE
An interactive user-interface has been designed as shown in Fig 1. The system will display a list of possible keyframes belonging to the highly ranked shots. These keyframes are displayed in an image display area on the left. A blown-up version of the cursor positioned image is shown on the right. The user can then label a shot as positive by clicking on it (selected images in red boxes). Additional shortcut keys are also available to enable effective communication between the users and the system.

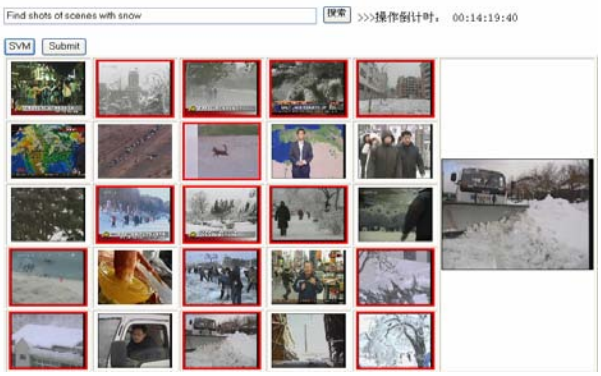

Figure 1. Interactive Interface

6. REFERENCES
[1] H.-B. Luan, S.-X. Lin, S. Tang, S.-Y. Neo, T.-S. Chua, "Interactive Spatial-temporal Visual Map Model for Web Video Retrieval", ICME 2007, Beijing, Jul 2007
[2] T. -S. Chua, S. -Y. Neo, Y. Zheng, H. -K. Goh, Y. Xiao, and M. Zhao, "TRECVID 2006 by NUS-I2R," TREC Video Retrieval Evaluation Online Proceedings, TRECVID 2006
[3] TRECVid, http://www-nlpir.nist.gov/projects/trecvid/
[4] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," In: Proc. ACM Intl. Conf. on Multimedia, pp.107-118. ACM Press, New York, NY (2001)

CIVR'07, July 9-11, 2007, Amsterdam, The Netherlands.