VisionGo: Towards True Interactivity
Yan-Tao Zheng, Shi-Yong Neo, Xiangyu Chen and Tat-Seng Chua
Lab for Media Search, Department of Computer Science, National University of Singapore, Singapore
{yantaozheng, neoshiyo, chenxiangyu, chuats}@comp.nus.edu.sg

ABSTRACT
This paper presents our interactive video retrieval system VisionGo. To achieve high interactivity, the system is fitted with a mouse stroke labeling interface, in which users can use mouse to draw the bounding box of target object via a simple mouse stroke. To enhance retrieval efficacy, the system leverages a new label class “near positive”, besides “positive” and “negative” label classes. This class of shots allows user to indicate shots that are incorrect but may be useful during feedback.

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

General Terms
Design, Experimentation

Keywords
Interactive Retrieval, Video Retrieval

1. INTRODUCTION
The target of interactive video retrieval is to bridge the semantic gap between raw signals from the video source and high-level semantic requirements of the searcher. The key lies in the interactivity between users and system. In other words, the knowledge from users’ annotation should be maximally utilized for effective retrieval, while their annotation efforts should be minimized for efficient labeling. To achieve such goal, our system leverages a mouse stroke labeling interface. This enables users to draw a mouse stroke across the target object, which results in a bounding box of object in the keyframe. Moreover, we also introduce a new label class “near positive” to assist the query expansion and relevance feedback process. Preliminary results from TRECVID 2008 [2] demonstrate the effectiveness and robustness.

2. USER INTERFACE (UI)
The system first provides the user with results from the automated search [1] for initial feedback. The user can make use of our intuitive UI to refine the search results with a variety of relevance feedback techniques, as show in Figure 1. To best leverage the annotation effort of users, we propose a mouse stroke annotation of target object region. A user can simply draw a mouse stroke across the object to obtain its bounding box, as shown in Figure 2. The subsequent visual process will be performed based on this object region. The rational is simple: compared to negative shots, the number of positive shots is quite small, which is worth spending more time on fine annotation.

3. TAGGING THE NEAR POSITIVE SHOTS
In most existing interactive retrieval systems in TRECVID and VideOlympics [3], the users are only required to provide one type of feedback, which is whether the shots are positive for the given query. The rest of the un-tag shots are automatically deemed as negative. To learn more knowledge from user’s annotation, we propose to add a new label class “near-positive”. This class denotes shots that are incorrect but may be useful during feedback. For example, the query “Find shots of a night scene with traffic lights”. This query requires two visual evidences of “night” or “traffic lights” to be both present for relevancy. However, if users tag a particular shot having only “night” or “traffic lights” scene as negative, this results in a loss of valuable annotation information. Even though these shots are incorrect with respect to the query, they can be useful in providing partial relevant semantics. In addition, it is highly probable that succeeding shots of a “night” shot may contain answer shots.

4. REFERENCES