ASPECT-BASED VIDEO BROWSING – A USER STUDY

Frank Hopfgartner, Thierry Urruty, David Hannah, Desmond Elliott, and Joemon M. Jose

Department of Computing Science
University of Glasgow
Glasgow G12 8RZ, UK

ABSTRACT

In this paper, we present a user study on a novel video search interface based on the concept of aspect browsing. We aim to confirm whether automatically suggesting new aspects can increase the performance of an aspect-based browser. The proposed strategy is to assist the user in exploratory video search by actively suggesting new query terms and video shots. We use a clustering technique to identify potential aspects and use the results to propose suggestions to the user to help them in their search task. We evaluate this approach by analysing the users' perception and by exploiting the log files.

1. INTRODUCTION

Current video retrieval approaches, in particular the retrieval systems evaluated in TRECVid [1], model retrieval in a "one result list only" approach, which assumes that the user is focused on one particular search issue. In a retrieval system, which provides the user with several independent result lists, one could search for information about various *aspects* of the underlying information need without interrupting the current search session. Within TREC-5 interactive track, the term "aspect" is used and defined as "roughly one of many possible answers to a question which the topic in effect posed" [2]. Similar topics were used in TREC-7 and TREC-8, indicating that retrieving different aspects is considered to be an important research question.

Harper and Kelly [3] use the aspectual search topics provided within TREC-8 to evaluate an information retrieval interface which provides the user with the facility to organise retrieval results within different *piles*. Each pile can then be used as a source of relevance information for executing new queries. Even though their study did not reveal a significant improving of retrieval performance, the participants of their study liked the approach.

Kerne et al. [4] introduce an interface which allows users to combine image and text summaries in order to promote idea generation and discovery. While providing a space for users to organise information, the focus of this system is more general however, not being solely focused on search tasks.

Villa et al. [5] propose an alternative video search environment by introducing a *faceted browser interface* which supports the creation of multiple search panels. Their study suggests that providing users with the facility to re-arrange retrieved results between panels eases their search session. Even though their approach is promising, further support is needed to help the users in their exploratory search.

A challenging question is how can users be assisted in the task of finding new aspects of a topic that they did not think about before. In this paper, we present a user study of an aspect-based video retrieval interface that automatically presents suggestions by extracting textual and visual features of selected relevant shots.

We aim to confirm whether automatically suggesting new aspects can increase the performance of an aspect-based browser. Therefore, we created a video retrieval system based on Villa et al. [5] and designed two different graphical user interfaces; a *Baseline interface* and an enhanced system, hereafter known as the *Suggestion* interface. The Suggestion interface provides the users with textual and visual suggestions, which are identified based on a clustering approach. We explore two main research questions:

1. Is an AspectBrowser a useful interface for aspect-based retrieval of videos?
2. Can textual or visual suggestions help users in finding new aspects of a topic?

The rest of the paper is structured as follows: we introduce our system in Section 2. Then, we briefly introduce the setting of our user study in Section 3 and analyse the resulting log files. Finally, we discuss the findings of our study in Section 4 and conclude in Section 5.

2. INTERFACE DESIGNS

2.1. Baseline Interface

Figure 1 presents a screenshot of the baseline interface, its annotated components are introduced in the remainder of this section. The interface is split into two vertical spaces, a topic space **(A)** and a retrieval space **(B)**. In the topic space, users can read their current search topic, including a textual description and the display of example images that show different aspects of the topic.

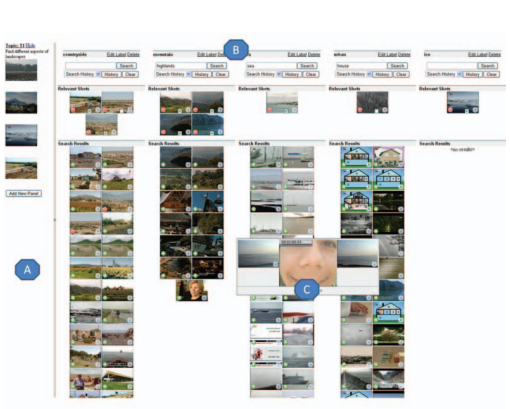

Fig. 1. Baseline

Fig. 2. Suggestion

Shots can be marked as relevant by dragging and dropping them to the relevant results list in a panel. Alternatively, they can be added to the relevant shots list by clicking on the green icon displayed in the left corner of each retrieved shot in the search results list. There is no restriction on which panel a shot can be dragged onto. Users can drag a shot from one panel directly to the relevant list of another panel, from an open video browser window **(C)** or from the topic panel, which allows the re-organisation of shots between panels. The drag action copies, rather than moves a shot so it remains in its source location after being dropped elsewhere.

Clicking on the small play icon of a keyframe will open a video browser window **(C)**. The shot represented by the clicked keyframe will start playing in the centre of the window, with keyframes of the shots temporally before and after displayed on the left and right of the video. Clicking on a keyframe will start playing the selected shot sequence and update the neighboured keyframes. This enables the user to browse temporally through a video, backward and forward.

Relevant rated keyframes contain a small tick box. Activating this box will nominate the selected keyframe as part of the next query-by-visual-example search query.

2.2. Suggestion Interface

The suggestion interface shown in Figure 2 is identical to the baseline interface, with the exception of the suggestion space (denoted **D**) positioned under the topic space. Figure 2 shows a collapsed topic space and the suggestion space. The suggestion space will update automatically every minute using the panel a user interacted with most during the preceding minute as the source of the suggestion. During the updating process, the background colour of the suggestion space changes to draw the user's attention to the update. As can be seen on Figure 2, the suggestion space consists of four main parts: the most frequent terms; keyframes based on most frequent terms; keyframes based on colour layout and edge histogram.

The text suggestion module is based on query expansion. A list of suggested keywords is generated based on the most frequent keywords which annotate the retrieved results in each search panel of our interface. These keywords are used to search for suggested video shots. The visual suggestion module is based on a clustering methodology. We make use of clustering to create groups of similar visual content. The clusters produced by our algorithm are assumed to be the aspects a real user may create in their search process. We believe a user's first query has a high probability of being general, with the retrieved set of results containing different semantic topics, e.g. if the query contains "sport" as keyword, the system will retrieve results of different sports and also other results such as people commenting on a match. Hence, we may obtain a set of more coherent aspects for the user, e.g. an aspect on "football" or "basketball" and another aspect on "people commentaries".

3. USER STUDY

In order to study the introduced research questions, we carried out a user study, which we describe in this section.

3.1. Experimental Design

We adopted a 2-searcher-by-2-topic Latin Square design where our 24 participants carried out two tasks using the baseline system and two tasks using the suggestion system. The participants indicated that they regularly interacted with and searched for multimedia. They were asked to search for each topic for a maximum of 15 minutes. Both the order of the questions and the order of the tasks were varied to avoid learning effects which could affect the outcome of the study. Each participant was given ten minutes of training on each system with a different training task for each system. The users' interactions with the system were logged and they were asked to fill out a number of questionnaires.

As suggested by Borlund [6], we created four simulated work task situations: "Find different aspects of landscapes", "Find different aspects of group work", "Find different aspects of education", and "Find different aspects of nautical vehicles". We chose the simulated work task situation to provide participants with a search situation to help them to better understand the task. We decided tasks which were not too complex to prevent the difficulty of the task interfering with our evaluation, as suggested by Bell and Ruthven [7]. The search tasks were designed based on the high-level feature extraction task within TRECVid 2008 to guarantee a satisfactory number of shots could be found in the collection. All tasks asked for different aspects of a broader topic and provided some examples. We manually picked example

Which system...	*B*	*S*	=
assisted in finding more relevant results?	4	14	6
was easier to learn to use?	10	2	12
was easier to use?	14	4	6
did you prefer?	9	12	3
changed your perception of the task?	5	5	14

Table 1. User Perception of both systems

keyframes showing different aspects of each task to provide participants with a starting point, if they needed one.

3.2. User Perception

After the participants finished using an interface to system, we asked them to rate the performance of the interface based on Five Point Likert scales. Some of these scales were inverted to reduce bias.

The participants agreed that the interfaces were effective in solving the search tasks. They also noted that the organisation of panels assisted them in analysing the task, in exploring the topics and in organising the shots they retrieved. They stated that the possibility to re-arrange results between search tasks resulted in a change of their usual search strategy. The most popular features mentioned were the possibility to carry out several searches at the same time, the different query facilities and the possibility to exchange results between different search panels.

Table 1 summarises the users' preferences between both systems. We were interested to find out which system was easier to use for the user. Most users stated that the baseline system was simpler to use and learn to use, which is not too surprising as it provides less visible features. Nevertheless, 50% of all users claimed that they experienced no difference in the difficulty to learn how to use both systems. The majority of our participants preferred the suggestion system over the baseline system. Negative remarks were that the suggestion panel distracted some participants within their search task. Some users felt out of control as the suggestions did not reveal what tab they were based on.

Our analysis of the participants' feedback suggests that they liked the idea introduced by the aspect browser. Moreover, they preferred the suggestion system, indicating that automatically displaying suggestions is a promising approach to improve their retrieval experience. In a next step, we evaluated their interactions with the interfaces to determine whether they used the provided suggestions.

3.3. Logfile Analysis

We analysed the interactions of participants to determine if their perceptions about the interfaces are supported by how they used the interfaces. We denote *Suggestion* as the proposed interface with the suggestion module and *Baseline* as the baseline interface. In our analysis, we consider a panel as a relevant aspect of the search process if at least one of the retrieved results has been dropped into the relevant results area of the panel. Moreover, we consider all keyframes being dropped into the relevant shots panel as being relevant.

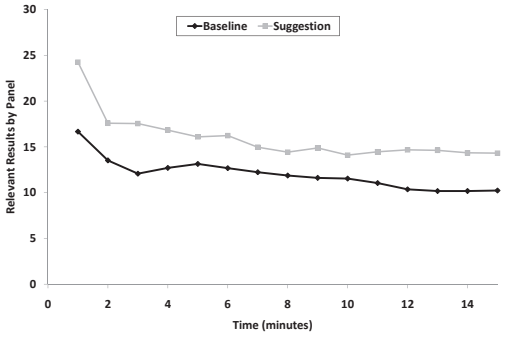

Fig. 3. The total no. of relevant results per panel over all topics

Figure 3 shows the average number of relevant results dragged and dropped in each panel. It clearly highlights that our proposed interface with a suggestion panel contains aspects created by users with more relevant results in each panel. The results above can be interpreted as users selecting more quality aspects with the *Suggestion* system.

	CL	EH	Text	Total
Actions	406	181	256	843
Actions (%)	48.16	21.47	30.37	100
Actions_{user}	16.92	7.54	10.67	35.12
$\text{Actions}_{user-topic}$	8.46	3.77	5.33	17.56

Table 2. Actions performed on suggestions

In order to analyse which type of suggestion is preferred by users, we compared the number of actions performed on the four suggestion types. For the most frequent terms suggestion, we identified only 7 keywords (*fans*, *people*, *masts*, *music*, *submarine*, *concerts* and *water*) that were added to the user query *after* being displayed in the suggestion space. These terms have a general meaning, thus it is not clear whether the user was influenced by their suggestion.

Table 2 present the actions performed on the suggestion panel. We count the number of 'play' and 'drag and drop' actions. The results show that users prefer the colour layout (denoted *CL*) based suggestions with 48.16% of the total number of actions, which is followed by the suggested shots based on text (denoted *Text*) with 30.37% and finally the suggestion from edge histogram (denoted *EH*) with 21.47%. We

also present the average number of actions by user and also by user and topic. We can conclude from the results in Table 2 that users do rely on suggestions quite often. Even so our results indicate a preference towards colour layout based suggestions, this observation might be biased by the fact that the position of the different types of suggestion are fixed (colour layout being in the middle of the suggestion space).

4. DISCUSSION

Our first research question was whether the *Suggestion* interface helps users to retrieve different aspects of the same topic. We did not have a predefined list of relevant results per topic so we rely on the participant's perception results to answer this question. The analysis of the questionnaires reveals that participants had a little preference for the *Suggestion* interfaces, as shown in Table 1. The questionnaires also asked whether participants would prefer to use an aspect-based interface for their daily multimedia searching and an average score of 4.5 out of 5 in their responses confirms that they would. So we can conclude that aspect browsing is a promising approach for effective video search.

The second research question was whether textual and visual suggestions could help users find new search aspects for a search task. The participants stated that the suggestion panel was an interesting idea as it was adapted to their information needs. The low number of textual suggestions which were subsequently used by participants from the suggested list of keywords indicates that they were not useful. This can partly be explained by the low quality of the provided text transcript which has been automatically translated from Dutch to English and a gap in the semantic meaning between the speech of the video and the visual content of the video.

The visual suggestion analysis showed that many actions were performed using them during each search session. Users did interact with suggestions, at least more than once per minute as shown in Table 2. It has helped users to identify new aspects. However, the analysis of the user's actions within different topics highlights several observations. Suggestions can be a distraction for the user in his search as stated by some participants in the open question area. A user cannot focus on selecting relevant results at the same time as interacting with the suggestions.

5. CONCLUSION

In this paper, we have presented an aspect-based video retrieval system with an automatic suggestion module. We proposed a user-centred evaluation of this system. This study was conducted on a large data set provided by TRECVid 2008. The study focused on evaluating whether automatic suggestions of terms or keyframes based on visual features could help the user in their search task. We were interested in discovering if aspect-based searching is an acceptable way for the user to retrieve their information needs. We studied our research questions based on system-centred and user-centred evaluation methodologies. The focus of our study was to determine which features were most useful for the user's search task. Users may not necessarily benefit from keyword-based suggestions, but they do use visual-based suggestions within the TRECVid setting, especially those provided by colour-based similarity.

This work is an entirely novel approach for video search and can address many deficiencies of current video search systems. We are investigating how this approach can be enhanced using an adaptive search model for video retrieval and how such an approach can be employed for practical video search.

6. ACKNOWLEDGEMENTS

This research was supported by the European Commission under contract FP6-027122-SALERO. It is the view of the authors but not necessarily the view of the community.

7. REFERENCES

[1] Alan F. Smeaton, Paul Over, and Wessel Kraaij, "Evaluation campaigns and trecvid," in *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, New York, NY, USA, 2006, pp. 321–330, ACM Press.

[2] Paul Over, "TREC-5 Interactive Track Report," in *The Seventh Text REtrieval Conference (TREC 5)*, 1999.

[3] David J. Harper and Diane Kelly, "Contextual relevance feedback," in *IIiX: Proceedings of the 1st international conference on Information interaction in context*, New York, NY, USA, 2006, pp. 129–137, ACM.

[4] Andruid Kerne, Eunyee Koh, Steven Smith, Hyun Choi, Ross Graeber, and Andrew Webb, "Promoting emergence in information discovery by representing collections with composition," in *C&C '07: Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition*, New York, NY, USA, 2007, pp. 117–126, ACM.

[5] Robert Villa, Nicholas Gildea, and Joemon M. Jose, "FacetBrowser: A User Interface for Complex Search Tasks," in *MM'08: International Conference on Multimedia, Vancouver, Canada*, 2008, pp. 489–498.

[6] Pia Borlund, "The IIR evaluation model: A framework for evaluation of interactive information retrieval systems," *Information Research*, vol. 8, no. 3, 2003.

[7] David J. Bell and Ian Ruthven, "Searchers' Assessments of Task Complexity for Web Searching," in *ECIR'04: Proceedings of the 26th European Conference on IR Research*, 2004, pp. 57–71.