



Urruty, T. and Hopfgartner, F and Hannah, D. and Elliott, D. and Jose, J.M. (2009) *Supporting aspect-based video browsing - analysis of a user study*. In: *Proceeding of the ACM International Conference on Image and Video Retrieval 2009*, Santorini, Fira, Greece, July 08 - 10, 2009. Association for Computing Machinery, New York, USA. ISBN 9781605584805

<http://eprints.gla.ac.uk/5757/>

Deposited on: 22 February 2010

Supporting Aspect-Based Video Browsing – Analysis of a User Study

Thierry Urruty
University of Lille 1
59655 Villeneuve d'Ascq
France
thierry.urruty@lifl.fr

Frank Hopfgartner
University of Glasgow
Glasgow, G12 8 RZ
United Kingdom
hopfgarf@dcs.gla.ac.uk

David Hannah
University of Glasgow
Glasgow, G12 8 RZ
United Kingdom
davidh@dcs.gla.ac.uk

Desmond Elliott
University of Glasgow
Glasgow, G12 8 RZ
United Kingdom
delliott@dcs.gla.ac.uk

Joemon M. Jose
University of Glasgow
Glasgow, G12 8 RZ
United Kingdom
jj@dcs.gla.ac.uk

ABSTRACT

In this paper, we present a novel video search interface based on the concept of aspect browsing. The proposed strategy is to assist the user in exploratory video search by actively suggesting new query terms and video shots. Our approach has the potential to narrow the “Semantic Gap” issue by allowing users to explore the data collection. First, we describe a clustering technique to identify potential aspects of a search. Then, we use the results to propose suggestions to the user to help them in their search task. Finally, we analyse this approach by exploiting the log files and the feedbacks of a user study.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces—*Evaluation/methodology*; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

1. INTRODUCTION

Current video retrieval approaches, in particular the retrieval systems evaluated in TRECVID [14] model retrieval use a “one result list only” approach, which assumes that the user is focused on one particular search issue. An example of this type of search task is: “Find shots of a vehicle approaching the camera”. These tasks are useful in benchmarking various retrieval algorithms as shown in TRECVID evaluation experiments, however, they are not representative of real world video information seeking tasks. A hypothetical example of such a task is: a journalist at a broadcasting station is searching for materials to produce an item for the evening news. He may be interested in highlighting the achievements of Michael Phelps at the 2008 Olympic Games

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIVR '09 July 8-10, 2009 Santorini, GR

Copyright 2009 ACM 978-1-60558-480-5/09/07 ...\$10.00.

in Beijing. However, as he progresses through the search task, he might be interested in highlighting the preparatory issues related to his performance, or in indicating the need for more governmental support for the development of swimming in the future. Current retrieval systems and approaches fail to provide any support for such parallel search activities.

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. 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.

The rest of the paper is structured as follows: we provide an overview of related work in Section 2. Further, we discuss our research questions in Section 3 and introduce our system in Section 4. Then, we introduce the setting of our user study in Section 5. Finally, we discuss the findings of our study in Section 6 and conclude in Section 7.

2. RELATED WORK

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” [11]. 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 improvement of retrieval performance, the participants of their study liked the approach.

Martin and Jose [8] introduce a system where users can group related web links into *bundles*. The system automatically analyses these bundles and proposes new search results.

Their study suggests that these automatically identified results match the users' current information need.

Kerne et al. [7] 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. [17] 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 big challenge is to find appropriate aspects, an important step towards further exploring the data collection. Users might not be aware of these aspects or might not know which queries to use to retrieve relevant results. Hopfgartner et al. [4] study this approach further; suggesting that a retrieval model which re-formulates search queries based on the content of parallel panels can increase the retrieval performance with respect to precision and recall. Furthermore, they introduce an approach for suggesting new aspects by clustering retrieved results based on textual and low-level features. They explore the performance bounds of an aspect-based retrieval system by using a simulated user evaluation methodology.

Aimed at evaluating this approach from a user perspective, we performed a user study on an aspect based video retrieval system. We compared two different graphical user interfaces; a *Baseline* interface and an enhanced system (hereafter known as the *Suggestion* system). The suggestion interface provides the users with textual and visual suggestions, which are identified based on a clustering approach.

3. RESEARCH QUESTIONS

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 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. Will the automatic suggestion of new aspects provide the user more relevant documents that they will interact with?
2. Can the suggestion system assist the user in finding more relevant results by relying on low-level features?

As we do not have a list of relevant results for these topics, we cannot analyse the system based on Precision and Recall, the standard evaluation measures in Cranfield-like evaluation approaches. Therefore, we evaluate our study by an interaction analysis. The system and its interfaces are described in the next section.

4. SYSTEM DESCRIPTION

The aspect-based browser is a web-based video retrieval system which is split into three main components: the retrieval backend that manages the visual and textual index of the data collection, the suggestion module and the graphical user interface. In this section, we describe the architecture of the system.

4.1 Data Indexing

The retrieval backend is responsible for the retrieval of relevant results based on search queries. Our system supports three types of search queries: Query-by-text, Query-by-visual-example and a combination of both types. To support these types of queries, we require two indices of our video collection: a visual index and a textual index. We merge the retrieval results from both indices as proposed by Wilkins et al. [18].

We use the TRECVID 2008 video collection as the collection of shots for retrieval. The corpus contains roughly 200 hours of various news documentaries and educational programmes from Dutch Television. The videos have been split into shots [12], the unit of retrieval used within TRECVID. A shot is a continuous recording of a video without break. Automatic speech recognition transcripts [5] and a machine translation from Dutch to English of all videos are provided. Moreover, roughly 20% of all shots do not have any textual annotation. The quality of the textual annotation hence is not adequate. We used the aceToolbox [9] to extract representative keyframes for each shot and extracted colour and edge histograms of these keyframes. The keyframes are then indexed based on these low-level features as suggested by Urruty et al. [16]. We removed stop words from the translated transcript, applied Porter stemming and indexed each shot using Terrier [10]. Okapi BM25 [13] is used to rank retrieval results.

4.2 Interface Designs

4.2.1 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. Clicking on "Hide" (**1**) will collapse the topic space and hide the example images.

New panels can be created by using the "Add New Panel" button (**2**) in the topic space. In Figure 1, a user created five search panels. Each search panel contains:

- A name for the panel, which can be provided by the user by clicking on "Edit Label" (**3**).
- A delete link, which removes the panel (**4**).
- The search box and button, allowing the user to enter a textual search and to trigger a retrieval (**5**).
- A pull down list of the searches already carried out in the panel (**6**). The user can re-execute previous queries by selecting an item on that list. The associated history button will pop-up a window containing deleted shots, which allows the user to undelete shots.
- The list of relevant shots as selected by the user (**7**).

- The list of search results (8), if any, ranked from most relevant to least relevant, top-left to bottom-right.

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.

Relevant shots can be removed from the relevance lists using the delete button on the bottom left of each keyframe. This will remove the shot from the panel and add it to the list of deleted shots, appearing in the history pop-up window.

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.

4.2.2 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. In addition, each retrieval panel contains an additional button labelled “Suggestion” (not shown in the figure) which will trigger the suggestion process using that panel as the source for the suggestion. During the updating process, the background colour of the suggestion space changes to draw the user’s attention to the update. Once the suggestion is displayed, the background colour changes to white again. As can be seen on Figure 2, the suggestion space consists of four main parts:

- Most frequent terms (9): the seven most frequent terms of each result list of the retrieval panel are displayed. Each line contains the terms extracted from the result list of one panel.
- Keyframes based on text (10): the seven most frequent terms of the according panel are used as search query and the ten top ranked results displayed.
- Keyframes based on colour (11): suggestions retrieved based on the colour layout of the input keyframes.
- Keyframes based on edges (12): suggestions retrieved based on the edge histogram of the input keyframes.

4.3 Suggestion Module

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”.

Our algorithm clusters the retrieved results based on low-level visual features. We choose to use agglomerative hierarchical clustering and the single link method [6]. Let C, D be two clusters, So_C and So_D are the respective set of objects of clusters C and D , the single linkage equation between C and D is given by the following formulae:

- for visual features of images representing video shots we use:

$$D_V(C, D) = \text{Min}\{d(i, j), \forall i \in So_C, \forall j \in So_D\} \quad (1)$$

where $d(i, j)$ is the Euclidean distance;

- for text queries, we use:

$$D_T(C, D) = \text{Max}\{d(i, j), \forall i \in So_C, \forall j \in So_D\} \quad (2)$$

where $d(i, j)$ is the number of common annotation keywords between two documents.

The output of the hierarchical clustering algorithm is a dendrogram. The number of clusters desired is an algorithm parameter, used to create the k clusters. We then create a new query for each cluster. As we use low-level visual features, we choose the medoid (the object closest to the centroid) of the cluster to create the new visual query. We assume the top k clusters form the k aspects of a user’s need and use them to create more specific queries. These queries will automatically propose new sets of results in the suggestion panel. Up to k result lists can be displayed as new suggestions.

5. USER STUDY

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

5.1 Experimental Design

We adopted a 2-searcher-by-2-topic Graeco-Latin Square design where our participants carried out two tasks using the baseline system and two tasks using the suggestion system. 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.

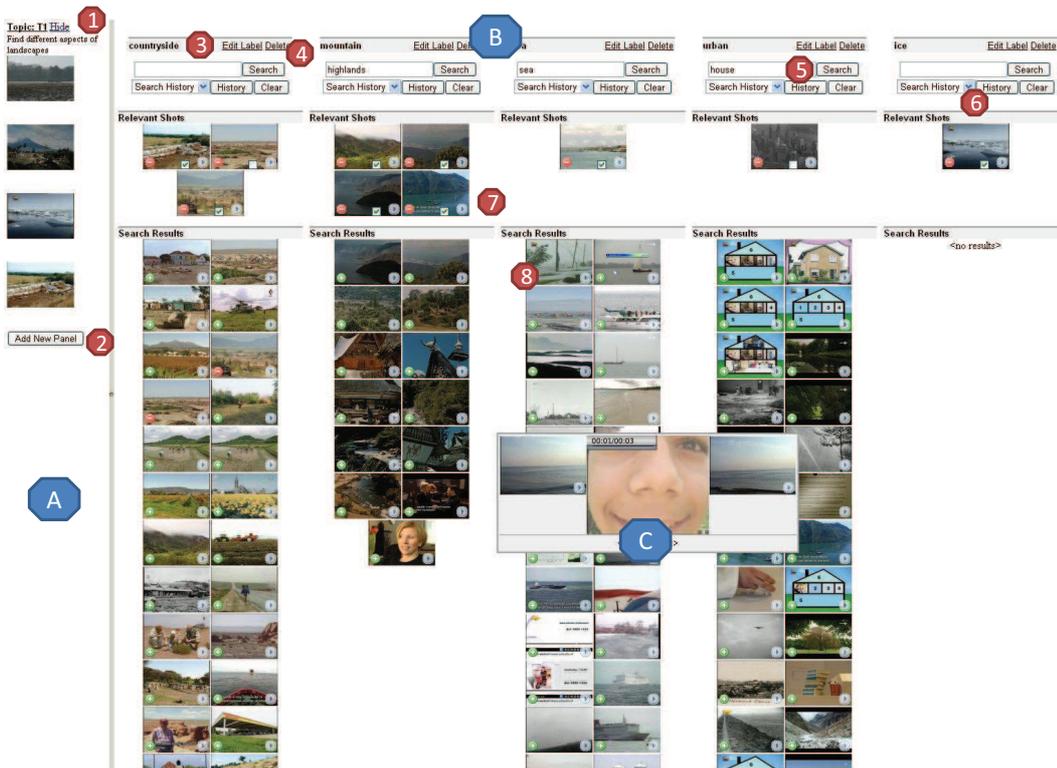


Figure 1: Baseline



Figure 2: Suggestion

The users' interactions with the system were logged and they were asked to fill out a number of questionnaires. The experiment started with an entry questionnaire, where users were asked to provide personal details and to rate their experience with multimedia. After each search task, they were asked to answer a post task questionnaire, aiming at understanding their opinion about the tasks. Then, after each two tasks, we provided them with a questionnaire about the system they used. Finally, an exit questionnaire was handed out where users were asked to compare both systems.

5.2 Participants

24 participants were paid £15 each to take part in our experiment. The participants were mostly university post-graduate students. The group of participants consisted of 18 males and 6 females with an average age of 27.6 years (median: 26). They indicated that they regularly interacted with and searched for multimedia and named YouTube as their favourite multimedia search engine. They claimed that searching for multimedia data is easy and they often find what they want. Their favourite search strategy when searching for different aspects of one topic was to rely on different tabs and using external sources (i.e. a web search) to identify relevant search terms.

5.3 Search Tasks

As suggested by Borlund [2], we created four simulated work task situations – these can be found in the Appendix. We chose the simulated work task situation to provide participants with a search situation to help them to better un-

derstand the task. We decided tasks which were too complex to prevent the difficulty of the task interfering with our evaluation, as suggested by Bell and Ruthven [1]. The search tasks were designed based on the high-level feature extraction task¹ 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 keyframes showing different aspects of each task to provide participants with a starting point, if they needed one. After each task, we asked the participants to evaluate the task based on a questionnaire containing Five Point Likert Scales. They all agreed that the task descriptions were clear and easy to understand. Moreover, they claimed that it was easy to think of different aspects covered by the tasks and that the video collection provided them with results of these different aspects. Overall, they felt comfortable searching for the tasks. A one-way analysis of variance (ANOVA) did not reveal a significant difference in the perception of the tasks ($p > 0.5$ for significance on average for all differentials).

5.4 User Perception

After the participants finished using an interface, 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. Table 1 shows the averaged differentials and the significant values based on a one-way ANOVA analysis

¹<http://www-nlpir.nist.gov/projects/tv2008/tv8.hlf.for.eval.txt.final>

given by the participants after using the baseline system B and the suggestion system S. The smaller the value, the more the users agreed with the differential.

The interface...	<i>B</i>	<i>S</i>	<i>p</i>
was globally effective	2.12	2.09	0.90
helped to analyse tasks	2.67	2.65	0.97
helped to organise documents	2.33	2.48	0.96
helped to organise the queries	2.50	2.00	0.32
helped to explore the topic	2.25	2.04	0.51
helped to express different aspects	2.13	2.35	0.51
helped to execute multiple searches	2.00	2.10	0.81
helped to reorganise material	2.54	2.09	0.20
was clumsy	3.67	3.43	0.52

Table 1: User Perception of the interfaces

The analysis did not reveal significant differences between the users’ perception of the interfaces which suggests any performance difference between the evaluated systems was solely caused by the suggestions provided.

As we were interested in the users’ perception of the provided suggestions, we asked them to answer further questions concerning the suggestion interface. The questionnaire revealed that our participants found the suggestions helpful to find more relevant videos. Moreover, the participants claimed that the suggestions were appropriate, indicating that the content of the suggested keyframes was relevant within this search task. They also stated that the suggestions helped them in identifying new aspects of the topic they did not think about before. Some quotations: “It gave some extra items which could be useful [...]”, “It gave suggestions difficult to obtain otherwise”, “[...] it helped in identifying other aspects of the results [...]”.

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 analysed their interactions with the interfaces to determine whether they used the provided suggestions or not.

5.5 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 present an analysis of all the search topics, and a detailed analysis of Task A and Task D because these reveal significant differences in interactions. 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. If a keyframe has been deleted from a panel, we remove it from the list of relevant rated results.

In a first step, we were interested to see if text played an important role in helping the users to find relevant results. Even though video transcripts might not be representative of the actual shot content [19]. Nevertheless, text can have a strong descriptive power. Table 2 shows the combined

	<i>Baseline</i>	<i>Suggestion</i>
Task A	182	143
Task B	184	126
Task C	160	160
Task D	144	178
Total	670	607

Table 2: No. of posed text queries

number of text queries that were used in the different search tasks. Even though a one-way analysis of variance (ANOVA) did not reveal significant difference between the two systems ($p > 0.45$), the total number of text queries is higher for the baseline system than for the suggestion system. Thus, we hypothesise that users had to rely more on text queries when using the baseline system, as the suggestion system automatically suggested shots based on low-level features.

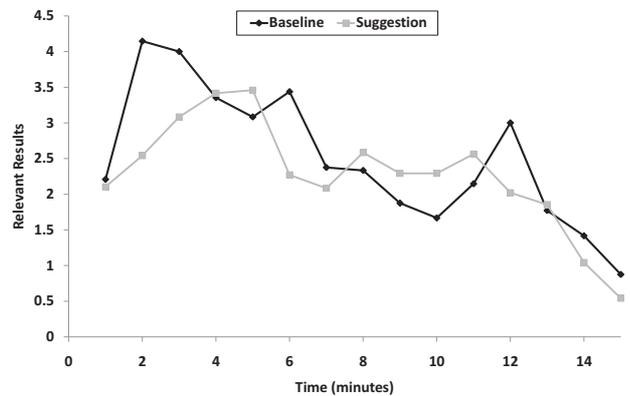


Figure 3: The number of average number of relevant results over all users and all topics

Figure 3 presents the average number of relevant results dragged and dropped by participants per minute. The overall tendency of both *Suggestion* and *Baseline* interfaces is a decrease in the number of relevant results with respect to time. This indicates that the user cannot find any more relevant results for the topic or that they have become demotivated or frustrated with the task. The main difference between both interfaces can be seen during the first four minutes. The *Baseline* interface peaks early and the *Suggestion* interface exhibits a reasonably steady rate of decline and increase. These results suggest the user takes time to consider the suggested results and explain the difference in total number of dragged and dropped results between the interfaces, presented in Table 3. This phenomenon has already been shown in the context of information retrieval. Initial stages of a user search session explore the collection with respect to the topic, then the user refines his search strategy and finally is making his relevant selection. Table 4 presents the number of panels created by users to represent a new relevant aspect of their searches. It suggests that participants create an average of 3.4 new aspects in the *Baseline* interface and 2.2 new aspects in the *Suggestion* interface.

Table 4 present the number of relevant panels created by users. It suggests that participants created on average 1

	<i>Baseline</i>	<i>Suggestion</i>
Task A	381	443
Task B	321	327
Task C	551	538
Task D	443	307
Total	1696	1615

Table 3: No. of relevant results

	<i>Baseline</i>	<i>Suggestion</i>
Task A	35	23
Task B	38	16
Task C	45	25
Task D	44	43
Total	162	107
Avg. by user	3.4	2.2

Table 4: No. of relevant panels

panel less with the *Suggestion* interface than with the *Baseline* system. This difference can be explained by the facts that the screen used could only display 3 panels without scrolling right or left and the width of the suggestion module on the screen is equivalent to one panel. Figure 4 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.

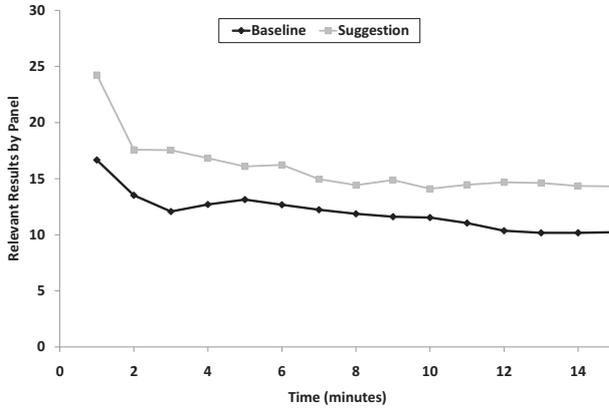


Figure 4: The total no. of relevant results per panel over all topics

The number of relevant results per minute varies with respect to the task. Table 3 shows that Task B and Task C produce similar results when comparing the *Baseline* and *Suggestion* interfaces. However, Task A and Task D show a noticeable difference in retrieval performance. Figure 5 shows the retrieval performance on average for Task A and Figure 6 shows the retrieval performance on average for Task D. Both figures compare the average number of relevant results dragged and dropped by participants during the search session. They depict two high peaks between both systems,

in favour of the *Suggestion* interface in Figure 5 and the *Baseline* interface in Figure 6. These peaks can be explained with the results presented in Figure 7 and Figure 8. These figures present the average number of relevant results for the *Suggestion* system alongside the average number of interactions with the suggestion module for Task A and Task D. For Topic A, Figure 7 clearly shows users are either interacting with the suggestion module or marking retrieved results as relevant. A high number of interactions with the suggestion module implies a high number of new relevant results will be marked in the next few minutes, so we can conclude that these new relevant results might have been suggested by our suggestions. For Task D, there is an unusual peak at the 12th minute in the *Baseline* interface, which can be explained by a participant who marked 14 shots as relevant in one minute (average of all other participants in 3 relevant results for the same minute) and then 0 as relevant for the rest of the task.

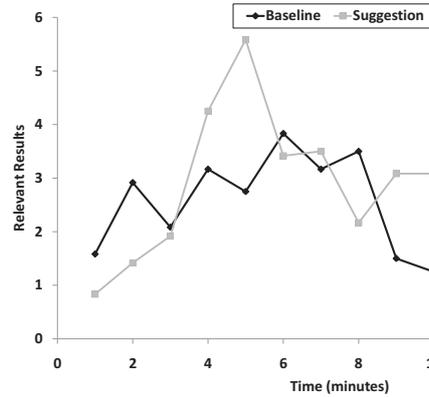


Figure 5: Average no. of relevant results per user for Task A

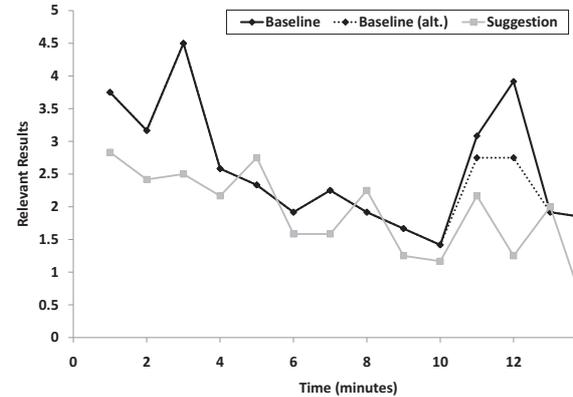


Figure 6: Average no. of relevant results per user for Task D

6. DISCUSSION

Our first research question was whether the *Suggestion* interface helps users to retrieve different aspects of the same

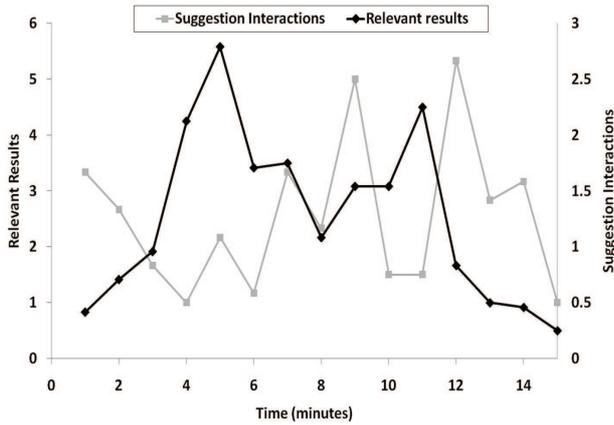


Figure 7: Average no. of relevant results wrt no. of suggestion actions for Task A

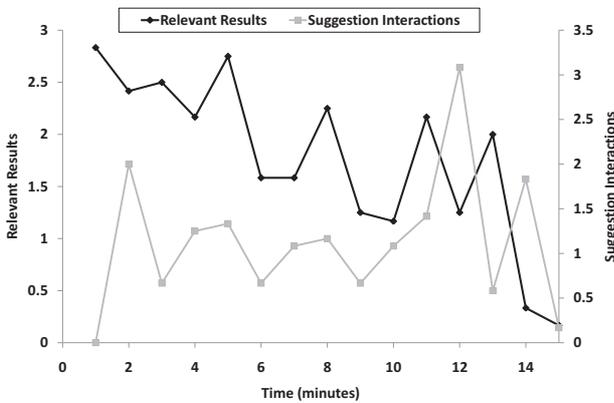


Figure 8: Average no. of relevant results wrt no. of suggestion actions for Task D

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 were equally satisfied with both the *Baseline* and *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 1.5 in their responses confirms that they would. So we can conclude that aspect browsing is a promising approach to 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 a interesting idea as it was adapted to their information needs. We noticed that users relied more on textual queries using the baseline system. We hypothesise that users had to rely more on text in this system, as the suggestion system provided them with additional shots they did not have to search for.

7. CONCLUSION

In this paper, we have presented an aspect-based video retrieval system with an automated suggestion module. This approach has the potential to address the “Semantic Gap” issue by allowing users to explore the data collection to a greater extend than normal search systems. We employed a clustering technique to identify potential aspects and used the results obtained from this technique to propose suggestions to the user to help them in their search task.

We presented 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. The analysis of the log files and the questionnaires seems to indicate the effectiveness of automatic aspect suggestion.

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.

8. ACKNOWLEDGMENT

This research was supported by the European Commission under contract FP6-027122-SALERO.

9. APPENDIX

9.1 Training Task A: Find different aspects of landscapes

Imagine you are working on a presentation of the different types of landscapes in the world for a class project. You want to give as broad a presentation as possible so you decide to use a collection of videos from an archive to help you.

Your task is to use the system to find different video shots of different landscapes, e.g. mountains, countryside, seaside, ocean, urban.

9.2 Training Task B: Find different aspects of group work

Imagine you are trying to understand different types of group work for a team project. Your team decides to create a presentation of the different types of group work to explain their understanding of the differences between working in different groups.

Your task is to use the system to find different video shots which show examples of people taking part in group activities, e.g. sports, playing music, and learning.

9.3 Task A: Find different aspects of education

Imagine you work for a TV content provider and are planning to make a documentary of different kind of education methods for different ages.

Your task is to find, using the system, different shots showing different aspects of education, e.g. primary school classes, university or adult teaching courses, private learning or sports courses.

9.4 Task B: Find different aspects of group gatherings

Imagine you are a police cadet and you have to give an overview to your colleagues about different situations they should be prepared for when large groups of people gather in their district. You decide to show your colleagues short video scenes as examples of previous events in your overview.

Your task is to use the system to find different video shots showing different types of gatherings of larger groups of people, e.g. in peaceful demonstrations, street riots, or parades.

9.5 Task C: Find different aspects of music sessions

Imagine you are the choreographer of a Bollywood movie and the director of the movie asked you to consider different music styles when directing different dance and music sequences for the various music scenes in his movie. Before providing the director with different music styles, you decide to study them using a collection of short videos.

Your task is to use the system to find different video shots showing different music scenes, e.g. when people sing alone, in a group, or dance. Also, you might want to distinguish between different music styles such as folk music, rock, or pop.

9.6 Task D: Find different aspects of nautical vehicles

Imagine you are the decorator of the Royal Nautical Museum and you want to create a short educational video about the variety of vehicles mankind has created to brave the dangers of the sea.

Your task is to use the system to find different video shots showing different nautical vehicles, e.g. canoes, submarines, aircraft carriers or rowboats.

10. REFERENCES

- [1] D. J. Bell and I. Ruthven. Searchers' Assessments of Task Complexity for Web Searching. In *ECIR'04: Proceedings of the 26th European Conference on IR Research*, pages 57–71, 2004.
- [2] P. Borlund. The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3), 2003.
- [3] D. J. Harper and D. Kelly. Contextual relevance feedback. In *IiX: Proceedings of the 1st international conference on Information interaction in context*, pages 129–137, New York, NY, USA, 2006. ACM.
- [4] F. Hopfgartner, T. Urruty, R. Villa, and J. M. Jose. Facet-based Browsing in Video Retrieval: A Simulation-based Evaluation. In *MMM'09: Proceedings of the 15th International Conference on Multimedia Modeling*, 01 2009.
- [5] M. Huijbregts, R. Ordelman, and F. de Jong. Annotation of Heterogeneous Multimedia Content Using Automatic Speech Recognition. In *SAMT'07: Proceedings of the Second International Conference on Semantic and Digital Media Technologies*, pages 78–90, 12 2007.
- [6] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, 1988.
- [7] A. Kerne, E. Koh, S. Smith, H. Choi, R. Graeber, and A. 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*, pages 117–126, New York, NY, USA, 2007. ACM.
- [8] I. Martin, and J. Jose. FETCH: A personalised information retrieval tool. In *RIAO '04: Proceedings of the 7th Computer-Assisted Information Retrieval (Recherche d'Information et ses Applications)*, pages 405–419, 2004.
- [9] N. O'Connor, E. Cooke, M. Blighe, and T. Adamek. The aceToolbox: Low-level audiovisual feature extraction for retrieval and classification. In *EWIMT'-5: Proceedings of the Second European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, pages 55–60, 2005.
- [10] I. Ounis, C. Lioma, C. Macdonald, and V. Plachouras. Research Directions in Terrier: a Search Engine for Advanced Retrieval on the Web. *Novatica/UPGRADE Special Issue on Web Information Access, Ricardo Baeza-Yates et al. (Eds), Invited Paper*, 2007.
- [11] P. Over. TREC-5 Interactive Track Report. In *The Seventh Text REtrieval Conference (TREC 5)*, 1999.
- [12] C. Petersohn. Fraunhofer HHI at TRECVID 2004: Shot Boundary Detection System. In *TRECVID 2004*, 2004.
- [13] S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. Okapi at TREC-3. In *Proceedings of the Third Text REtrieval Conference (TREC 1994)*, Gaithersburg, USA, 1994.
- [14] A. F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and trecvid. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press.
- [15] K. Spärck-Jones and C. J. van Rijsbergen. Report on the need for and provision of an ideal information retrieval test collection. Technical report, University Computer Laboratory, Cambridge, 1975. British Library Research and Development Report 5266.
- [16] T. Urruty, C. Djeraba, and J. M. Jose. An efficient indexing structure for multimedia data. In *Proceedings of ACM MIR 08*, Vancouver, Canada, 2008. ACM.
- [17] R. Villa, N. Gildea, and J. M. Jose. Facetbrowser: a user interface for complex search tasks. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, pages 489–498, New York, NY, USA, 2008. ACM.
- [18] P. Wilkins, P. Ferguson, and A. F. Smeaton. Using Score Distributions for Query-time Fusion in Multimedia Retrieval. In *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia Information Retrieval*, pages 51–60, New York, NY, USA, 2006. ACM.
- [19] R. Yan and A. G. Hauptmann. Co-retrieval: A boosted reranking approach for video retrieval. In *CIVR*, pages 60–69, 2004.