

# Differences in Video Search Behavior between Novices and Archivists

Henning Rode, Theodora Tsikrika, and Arjen P. de Vries

henning.rode@cwi.nl   theodora.tsikrika@acm.org   arjen@acm.org  
CWI Amsterdam, The Netherlands

**Abstract.** Improving the user’s interaction with a video retrieval system requires to examine the search behavior of real users. We present in this article a study comparing the video search behavior between professional archivists and novice users. The comparison focuses on the use and effectiveness of different state-of-the-art video search methods offered by our retrieval system, and the result investigation behavior of the two user groups. We conducted our experiments in the context of TRECVID’s 2009 interactive search task, using the provided collection and topics for our evaluation. The findings are based on a qualitative questionnaire analysis and a quantitative examination of the logged user actions on the search interface. The experimental results indicate that today’s visual search techniques have improved in effectiveness, confirming a trend found in previous user studies. To our surprise, professional archivists used visual concept search in many of their searches. Queries containing visual concepts were more effective, resulting in more relevant shots found than the alternative methods. Overall, we conclude that professional archivists are more focused on recall in carrying out their search tasks, and are better at reflecting on their own search performance.

## 1 Introduction

VITALAS is an EU-funded Integrated Project that aimed at the development of a system capable of large-scale indexing and retrieval of video and images, specifically targeted towards multimedia professionals and archivists. The VITALAS video retrieval system integrates different state-of-the-art search methods into a single user interface that supports search on (the combination of) multiple modalities. This paper reports on two series of experiments, the first carried out within the context of TRECVID’s 2009 interactive search task, and a later independent follow-up experiment that used the same test collection. The aim of both experiments has been to analyze and compare the use and user-perceived effectiveness of the available search functionalities. We examined how different types of users interact with the system, by comparing the behavior of professionals and novice users. Archivists employed by three large-scale broadcasting archives (located in France, Germany and The Netherlands) participated as professional users in our study, while novice users were represented by people who

are not professionally concerned with video retrieval. We performed a quantitative and qualitative evaluation of users' search behavior by using questionnaires and analyzing the search logs.

The remainder of this paper is structured as follows. Section 2 presents related user studies that examined the behavior of users interacting with video retrieval systems. Section 3 describes the VITALAS video retrieval system, its search components and user interface, and the used test collection. Section 4 details the set-up and execution of the user tests. Section 5 presents the results of the experiments and tries to point out differences and common findings with the above mentioned related user studies. Section 6 concludes this work.

## 2 Related work

Several previous works studied multimedia retrieval in an interactive setting. For example, Westman et al. conducted a comparison of professional and non-professional users of an image retrieval system [15]. Editors of the magazine industry participated as professionals in a study where they carry out retrieval tasks representative of their daily work. Hollink et al. analyzed the search patterns arising from the use of their news video retrieval system by non-professional users [11], distinguishing search behavior with respect to different search task categories; contrasting searching for specific known items or entities with more generic search tasks.

The majority of user studies that have examined the video searching behavior of users interacting with video retrieval systems have however been carried out in the context of TRECVID's interactive search task, see e.g. [14, 13] and the studies discussed in more detail below. In the TRECVID framework, interactive search is defined relative to a given video collection, a common shot reference for this collection, and a multimedia statement of an information need referred to as *topic*, with the aim to return a ranked list of up to  $N$  ( $= 1000$ ) shots which best satisfy the need. Users participating in the task have no prior knowledge of the search test collection or topics. A review of the video retrieval approaches investigated during the first three years of TRECVID (2001-2003) revealed that interactive approaches consistently and substantially outperformed all non-interactive approaches [9]. In this period, "query-by-text", rather than "query-by-image" and "query-by-concept" accounted for most of the successful interactions, particularly among novices [5].

Our study is most similar in goals and research methods used to the research carried out by the Informedia research group (during several years of TRECVID), so we will discuss their work in more detail [3, 4, 2, 6]. Similar to our analysis, these studies combine questionnaires and search log analysis in order to examine user behavior. They compare in depth the effectiveness of different search methods and querying paradigms, looking separately into the performance of "expert" users and "novices". Here, the authors distinguish experts from novices on the basis of three types of knowledge that only their experts possess [2]: (1) experts have been working with the research group that develops

the video retrieval system to be evaluated, thus having a better understanding of the various automated video processing techniques; (2) the expert has used the tested video retrieval system prior to the user study with the TRECVID data, perhaps even contributing to its development, and therefore knows the system operation better than the users who first interact with it during the user tests; and (3) the expert is familiar with TRECVID evaluation, e.g., the emphasis on shot-based retrieval and use of mean average precision as a key metric. (According to this definition, our professional archivists would not be considered experts!) In their TRECVID 2004 user studies [3], it was observed that expert users were more willing to use image-based and concept-based search, whereas novice users mostly engaged in “query-by-text” interactions. The much better performance by the experts in that year’s search tasks was attributed to the use of all three querying strategies and motivated the development of a video retrieval system that would enable increased use of the different search methods. The Informedia interface was re-designed so as to promote the use of non-text based video access mechanisms [4], achieving the desired effect that novice users increased the use of image-based and concept-based searches, rather than exclusively relying on text search. In TRECVID 2006 [2, 6], the “query-by-text” strategy was even less dominant and the employment of concept-based queries was observed to be an effective strategy for reducing the result shot space. In both 2005 and 2006, the Informedia team had government intelligence analysts participate in the experiments.

While these studies share the overall aims and research method with the work discussed here, notable differences in setup justify the effort put into our own experiments. First, the TRECVID collection has changed considerably, changing its contents from the broadcast news domain to a more heterogenous collection that includes documentaries and educational programming. TRECVID’s 2009 search topics concentrate on generic search tasks only, not including known item or entity searches. Also, previous user studies were conducted more than two years ago, a time span in which especially the concept-based retrieval methods improved according to the results obtained in fully automatic machine-only evaluations. Finally, we believe the archivists participating in our study represent a different type of professional users that come with different experiences and needs, even when compared to intelligence analysts.

### 3 Video Retrieval System and Test Collection

The VITALAS video retrieval system used in this study allows users to make use of the following search functionalities: (1) keyword search, (2) concept search, (3) visual similarity search, (4) fused search (any combination of the above), and (5) concept suggestions.

The text search component allows search on the text output generated from automatic speech recognition (ASR) on the video material. It provides common full-text search functionalities, such as keyword and phrase search, and returns a ranked list of shots. The concept search retrieves shots based on automati-

cally detected concepts. Similarly to keyword queries, the user can search for any combination of concepts and the system ranks the shots in the collection based on the estimated combined relevance. Visual similarity search completes the query functionalities offered. Once users have found one or more relevant examples, they can use them to search for shots in order to find visually similar keyframes. Similarity search thus enables the retrieval of shots based on visual features, without being bound to the predefined set of concepts. Users are further supported by a concept suggestion service. Whenever a user issues a text search, the service returns concept suggestions related to the submitted query. In this way, users are made aware of automatically detected concepts that might be useful for refining or expanding their queries.

Figure 1 shows the main query interface as well as the *result view*. The top text field allows users to enter keyword and concept queries. The retrieved shots are shown by a thumbnail keyframe in the mosaic result view below. Each thumbnail can be added to a lightbox that is used for gathering possibly relevant shots for the search topic. Clicking on thumbnails opens a *zoom view* showing a single keyframe in a higher resolution as presented in the figure. The zoom view also enables to enter a *detailed view* in order to play the shot (Figure 2). The thumbnails can also be used to issue a visual similarity search. The detailed view allows users to play the shot and to jump to any other shot in the video. (At the time of the experiments, the test prototype of the system did unfortunately not allow to add other shots from the detailed view to the lightbox.) Hotkeys allow a more efficient user handling for the experienced user. The “suggestion bar” shows concept suggestions derived from the previous query’s results, that upon the user’s click are added to the current query.

The TRECVID test collection for the search task in 2009 consists of 280 hours of Dutch video data of news magazines, science news, news reports, documentaries, educational programming, and archival video from the Dutch national broadcasting archive. For the keyword search component, we used a machine-translated version of the ASR output provided by the LIMSI system [8]. (We

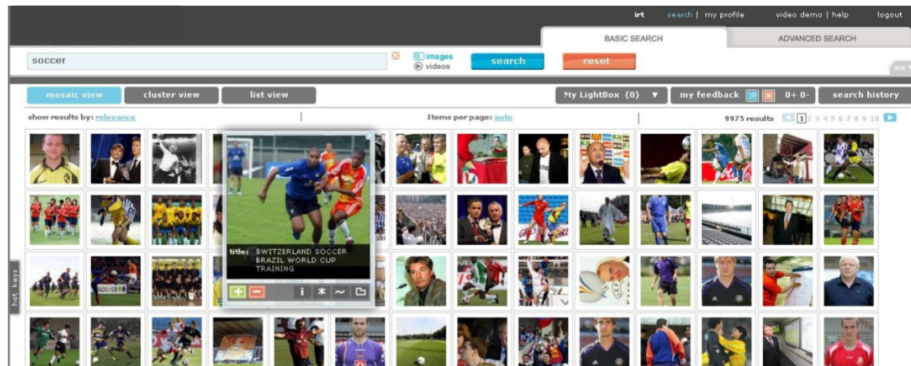


Fig. 1. VITALAS Video retrieval system user interface: zoom view

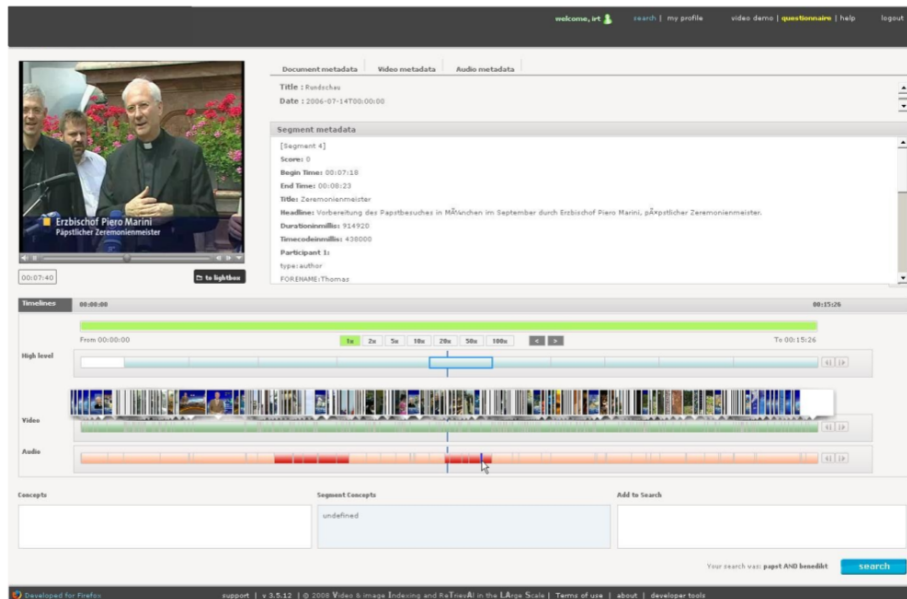


Fig. 2. VITALAS Video retrieval system user interface: detailed view

used the Google machine translation services<sup>1</sup> to automatically translate the speech transcripts to English text.) The collection is also annotated with 64 different concepts detected automatically. We used the concept detector output for the 20 TRECVID 2009 concepts generated from CERTH-ITI's concept detector described in [7], extended with the detector output for an additional 44 concepts publicly provided by the MediaMill<sup>2</sup> [13]. Machine-translated transcripts and concepts were indexed and made searchable using the open source PF/Tijah retrieval system [10]. For the visual similarity search, features of all keyframes were extracted and indexed by Maestro, INRIA's visual search component described in [1].

## 4 User Test Setup

The user tests were split in two phases, one taking place in the context of TRECVID's 2009 interactive search task and one in a later independent user study that also employed the same test collection. The system deployed in the second test phase was an upgraded version of the TRECVID prototype. Due to the system differences, albeit rather minor as explained below, the two test phases are examined separately without conflating the user groups that partici-

<sup>1</sup> <http://code.google.com/p/google-api-translate-java>

<sup>2</sup> <http://www.science.uva.nl/research/mediamill/>

pated in each of them. It should be noted that none of the test users participated in both test phases.

#### 4.1 Test Phase 1

The first test phase was conducted by following the schedule of the TRECVID 2009 interactive search task. We recruited a total of 10 users to participate in our experiments: 4 professional archivists employed in institutes hosting large archives of public broadcasting (3 from France and 1 from The Netherlands) and 6 non-professional users. None of the users have been involved in the design or implementation of the VITALAS video retrieval system; therefore, in order to gain some familiarity with the system interface and supported functionalities, all users completed a training session prior to their main search sessions. Given their familiarity with the daily use of thesauri for searching their own archive, professional users were additionally provided with a list of the available concepts. Both user groups could use the system's concept suggestions to modify their searches.

Each user was required to complete 12 of the 24 TRECVID 2009 topics, assigned to them based on a latin squares arrangement. (The order of the topics was not randomized, so that in principle a learning effect across user groups could be observed.) Each user could spend a maximum of 10 minutes on each topic before proceeding to the next one. Users were instructed to save those shots that they considered to be relevant to the topic in question. However, the instructions did not emphasize that they should find as many relevant shots as possible, which resulted in only a few saved shots per topic (about 9 on average). They may have focused on only the shots they considered highly relevant, instead of the actual TRECVID task of collecting all relevant shots. The system logged all user interactions, including the submitted queries, the shots viewed, and the shots selected as relevant (i.e., added to the lightbox).

Apart from collecting the raw user-interaction data, users were asked to fill in questionnaires at different stages of the experiment: (i) an *entry questionnaire* for collecting background information on the searchers, (ii) a *search questionnaire* provided after each topic to ask users about their perception of the just performed search, and (iii) an *exit questionnaire* which asked for an overall evaluation of the VITALAS system and the functionalities it offers.

#### 4.2 Test Phase 2

A second phase of the user tests was conducted half a year later, when a different group of professional archivists was available. Meanwhile, the retrieval system had undergone a few changes - an unavoidable drawback of carrying out this research using an integrated system consisting of components that are being developed independently.

The largest difference between the two system variants is that users in the second phase could see zoom views of keyframes by mouse-over, instead of requiring a click on the thumbnail representation in the result overview. This

modification eased the interaction with the system, and consequently increased the number of zoom actions in the searchlogs; but, at the same time, each zoom interaction being less of a conscious user decision. A second, rather unfortunate difference between the two system variants is that the concept suggestion service was not available during this second test period. As a resolution, all users (professionals as well as novices) were provided with the list of automatically detected concepts.

Apart from these system changes, topic assignment had to be modified as well, as the second group of users was more time constrained with respect to total availability for the entire test. We decided to stick to the maximum duration per topic of 10 minutes, and therefore had to reduce the number of topics. We had 4 archivists (located in Germany), each working on 8 TRECVID 2009 topics, and 5 novice users completing 4 topics each. Similar to the first test phase, all users were asked to fill in questionnaires, and all search interactions were logged by the system.

## 5 Results

This section presents results from evaluating the gathered questionnaire data and complements the qualitative information from the questionnaires with a quantitative analysis based on the interaction data collected during the experiments.

### 5.1 Questionnaire-based Analysis

Table 1 presents users’ own perception of effectiveness and ease of use of the available search methods, averaging the assessments collected in their exit questionnaires.

**Table 1.** Users’ perception of search methods.

		archivists		novices	
test phase	search method	ease	effectiveness	ease	effectiveness
phase 1	text	4.25	3.00	<b>4.00</b>	3.16
	concept	4.25	<b>3.50</b>	3.00	3.50
	similarity	4.25	2.00	3.16	3.00
	fused	4.25	2.75	3.66	<b>3.83</b>
phase 2	text	3.66	2.25	<b>4.60</b>	3.00
	concept	4.00	<b>2.75</b>	3.40	<b>3.50</b>
	similarity	<b>4.25</b>	1.25	4.00	3.20
	fused	3.00	2.25	3.20	3.00
questionnaire scale: 1 – 5					

Archivists indicated no difficulties in using the new, visual access methods (though not available in their usual daily work practice). Novice users expressed

a preference for the text search over the other search methods, with respect to ease-of-use. Both groups value the concept search highest with respect to the perceived search effectiveness, followed by fused and text queries. In general, novice users tend to give higher grades with respect to the perceived effectiveness of the different search methods. It should be remarked that the table only shows averages. Differences between individual users are rather large, resulting in a high variance of individual values, in particular for the novices (in line with findings reported by [16], further discussed in the conclusions of this paper).

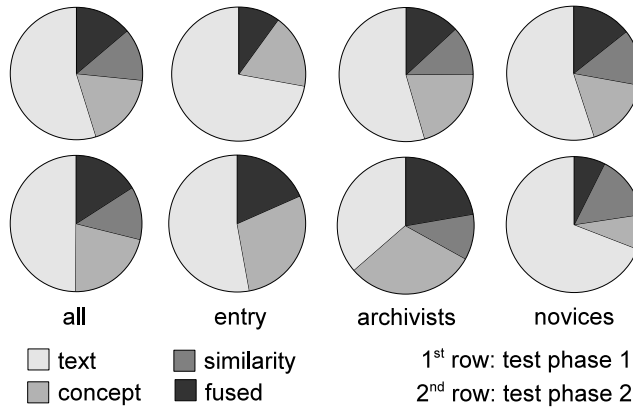
The search questionnaires inquired after each topic assignment whether the user felt they had had sufficient time to work on the topic. While novice users judged the given 10 minutes as sufficient in all cases, independent from the actual perceived search success, professionals would have liked to continue the search on a number of topics, in those cases where they expected to find more relevant results. We also observe for professional users a higher correlation between the self-judged completeness of a search and the satisfaction with search time. The Informedia study [2] does not compare novices and professionals in this respect, but our findings for the professionals confirm their results: the intelligence analysts in their studies were also willing to search longer, even when the satisfaction with the found results remained low. Novices expressed a higher confidence in the quality of their search results, although their self-assessment of success will not be confirmed when we later take a look at the effectiveness measured by the TRECVID assessments.

## 5.2 Search-log Analysis

The analysis of interaction data concentrated on the following four aspects of search behavior: (1) the use of the available search methods, (2) the effectiveness of these search methods, (3) the agreement between TRECVID assessors and our test users, and (4) users' interaction with the displayed results.

**Use of the Search Methods** Figure 3 shows how often users made use of each search method. More precisely, we count the number of queries that contain at least one text search, concept search, similarity search, or fusion predicate. Therefore, a single query can count for multiple methods at the same time. We observe that the text search was employed most often, followed by concept search, fused search, and similarity search. Looking only at the first query that a user issued when starting with a new topic (referred to as *entry* search), we can see that text and concept search are most often used as entry queries. Similarity searches could not be used to start a new topic since the system only allows to start a similarity search from already found keyframes.

During the first test phase the distribution of search methods stayed similar among the user groups. However, we can observe a clear difference between professional and novice users in the second test phase. Especially the concept search was more widely employed by the professionals. This could be partly attributed to the fact that concept suggestions were not provided during the



**Fig. 3.** Proportions of the different search methods.

second test phase. Apparently, professionals were nevertheless willing to use concepts by consulting the provided list of available concepts, while novice users concentrated on other search methods.

A comparison with the Informedia study [2] shows a number of similarities: (1) Their professionals also made more use of the visual search methods than the novices. (2) Among the visual search methods, the novices slightly prefer the similarity search, while professionals use the concept search more often. The tendency of more visual searches is confirmed when taking also the slightly older study of Hollink et al. [11] into account. Of course, we should be careful drawing firm conclusions from comparing these studies with ours, as those were carried out on a different data set. Both previous studies report an average number of issued queries per topic of approximately 7 query reformulations, with a lower proportion of visual search methods applied. While we cannot tell why our user groups issue visual search methods more often, we identified three possible (partial) explanations: the improvements observed in automatic benchmarks of visual search techniques, a relatively low effectiveness of searching machine-translated speech transcripts, and the larger heterogeneity in the collection.

**Effectiveness of the Search Methods** With respect to the effectiveness of the different search methods, the analysis can be performed by a system-oriented or a user-oriented perspective. A system-oriented analysis examines the number of results retrieved by the system that are also judged as relevant by the TRECVID assessors. A user-oriented analysis considers the number of results added to the lightbox by the test users and thus considered as relevant by them. Table 2 presents the results of these two analyses. Concept search is by far the most effective search strategy from the system perspective. The high difference may be explained by the fact that most TRECVID topics have well matching available concepts. From a user perspective, the differences are considerably smaller, but still a search containing concepts leads on average to a higher number of shots

added to the lightbox. Hence, we can state that the measured system level effectiveness is confirmed by the user experience. The results also roughly correspond to the user experiences expressed in the questionnaires shown in Table 1. Hence, the users were able to estimate the effectiveness of the different methods. We cannot explain why testers of the second phase selected in all cases less shots to add to the lightbox, but at least the results of the two test phases confirm each other with respect to the order of effectiveness.

**Table 2.** System vs. user effectiveness.

search type	test phase 1		test phase 2	
	relevant retrieved	added to lightbox	relevant retrieved	added to lightbox
text	2.76	0.98	3.53	0.74
concept similarity	<b>13.47</b>	<b>2.69</b>	<b>12.48</b>	<b>1.04</b>
fused	4.32	1.07	3.24	0.63
	7.88	1.79	7.2	0.70

**Agreement with TRECVID assessors** We also evaluated the agreement between our test users and the TRECVID assessors. We estimate the agreement level here using two measurements: (1) we look at shots selected by the user and added to the lightbox, and compute the ratio of agreement with the assessors, and (2) we look at shots that users have examined in more detail by zooming in, but have not been added to the lightbox. We interpret the latter shots as consciously rejected by the user. Table 3 displays the results of this analysis. The first test phase revealed a high level of disagreement, which can however be attributed to a rather high number of user judged shots that was unjudged (i.e., were not part of the assessment pool). In the second phase we observe a considerably higher agreement of about 75% for shots added to the lightbox, and almost 70% for zoomed-in shots. Such agreement level seems more reasonable, and comparable to the agreement measured in the Informedia user study [2].

**Comparison between users** We also examined whether users who had been assigned the same topic found the same or different shots. Although each topic

**Table 3.** Agreement between users and TRECVID assessors.

	test phase 1		test phase 2	
	relevant	irrelevant	relevant	irrelevant
added	273	251	106	35
rejected	150	196	348	745
total	423	447	454	780

was assigned to a total of five users, the proportion of common shots found by more than one user within all added shots for a topic is only 17%. Hence, by far most of the added shots for a given topic are unique among our users. This could be due to the fact that our users added on average only 9 shots per topic to the task search results. This number is even lower than the 30 shots per topic found by the Informedia users [2] (who could however search for 15 minutes each). We agree therefore with Christel’s assumption that real users are not willing to search for arbitrary many relevant results; after finding several relevant shots, users regard the topic as finished.

**Result Investigation Behavior** A further analysis examined the interaction of the users with the displayed system results. Especially, we looked at how often users zoomed in on a displayed keyframe in order to judge its relevance – zooming in on the keyframe and watching the video were both considered as zoom actions – and how often they selected shots as relevant for their search; this is different from shots being judged as *relevant* by the assessors. We mention here the results of the first test phase only and later discuss the differences to the second phase.

Our results indicate that both user groups perform almost the same number of zoom actions. Professional users investigate however shots much deeper in the ranked retrieved list, as the median rank of their zoomed and selected items is twice as high when compared to novice users. On average, the total number of zoom actions of users (irrespective of the user group they belong to) is twice as high as the number of select actions, which indicates that the initial result overview showing thumbnail keyframes is often insufficient to judge a shot on relevance.

Looking at the results of the second test phase, we observe the same difference between professional and novice users described above. The median rank of selected items for the archivists lies at 38, respectively 19 for novices. Compared to the first phase we further see a highly increased number of zoom actions due to the new mouse-over trigger in the search interface. In total the search logs of the second phase contained 4682 zoom actions, 317 select actions, but only 103 times users opened the detail view in order to watch the video shot. Hence, most shots were selected based on their keyframe rather than by watching the video. These numbers are, however, highly determined by the search interface. If the interface could also lower the burden to watch a shot as it was done for zooming the keyframes, we may expect more users to view a shot before selecting it as relevant.

## 6 Conclusions

In 2009, Wilkins et al. [16] presented an analysis of a multi-site video retrieval experiment using the TRECVID framework, where they discovered that non-expert users generated very large performance fluctuations. We find a high variance in our experimental data as well, especially among the novice users. However, the

primary purpose of our experiments is not a direct comparison between the performance of different video retrieval techniques, but rather identify differences in search behavior between the two user groups. While we cannot conclude that one technique performs consistently better than another one, we *can* draw conclusions such as “professional users apply concept-based searches more frequently”, or, “novices are more likely to consider their information needs satisfied in spite of low recall”. Irrespective of the large variance between individual users in each group, clear patterns arise from observing their aggregated observations.

The evaluation of the different search methods used showed clearly that simple text search is employed as the default method, especially when starting a new search, even if other methods can be more effective, and, are also experienced as such. We expect that more training to gain a better knowledge of all available concepts, improved concept suggestion functionality, and, the possibility to start a search by visual similarity initiated by an uploaded image (e.g., as the result of a web image search), are three directions that can help users make more effective use of the visual search techniques. One could argue that the observations in our data with respect to the effectiveness of visual search techniques are in line with the findings of recent evaluations of automatic systems at TRECVID. Users in our experiments value the “query-by-concept” and “query-by-example” search methods more highly than those reported in earlier experiments.

The differences in search behavior of the two user groups in our study are most clearly visible when looking at the result investigation behavior. Professionals are willing to spend much more time to investigate the presented results, especially when this behavior leads to more and better results. A search system designed for professional archivists should therefore be optimized for high recall, and enable the user to easily investigate the relevance of found shots. For the VITALAS system, two straightforward interface improvements would be to let the user easily modify the size of shown keyframes, and to provide the possibility to play a shot by mouse-over (instead of the two clicks required now).

Although we tried to point out the importance of user studies conducted with real users – instead of the often presented results from experts being involved with the system development – it is at the same time undesirable for a realistic user study that professionals are not familiar with search system they have to use. While a short training session could overcome this problem partially, we anxiously await the time when we can repeat this study with professional archivists who have actually used visual search techniques in their daily work practice.

## 7 Acknowledgements

This paper was realised in part by funding from the European Union via the European Commission project Vitalas, contract no. 045389. We are grateful to our colleagues from INA, IRT and Beeld & Geluid, for arranging the tests involving archivists on their premises.

## References

1. N. Boujemaa, J. Fauqueur, M. Ferecatu, F. Fleuret, V. Gouet, B. Le Saux, and H. Sahbi. Ikona: Interactive generic and specific image retrieval. In *Proceedings of the International Workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR 2001)*, 2001.
2. M. G. Christel. Establishing the utility of non-text search for news video retrieval with real world users. In *Proceedings of the 15th ACM International Conference on Multimedia*, pages 707–716, New York, NY, USA, 2007. ACM.
3. M. G. Christel and R. M. Conescu. Addressing the challenge of visual information access from digital image and video libraries. In M. Marilino, T. Sumner, and F. M. S. III, editors, *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL 2005)*, pages 69–78. ACM, 2005.
4. M. G. Christel and R. M. Conescu. Mining novice user activity with TRECVID interactive retrieval tasks. In H. Sundaram, M. R. Naphade, J. R. Smith, and Y. Rui, editors, *Proceedings of the 5th International Conference on Image and Video Retrieval (CIVR 2006)*, volume 4071 of *Lecture Notes in Computer Science*, pages 21–30. Springer, 2006.
5. M. G. Christel and N. Moraveji. Finding the right shots: assessing usability and performance of a digital video library interface. In Schulzrinne et al. [12], pages 732–739.
6. M. G. Christel and R. Yan. Merging storyboard strategies and automatic retrieval for improving interactive video search. In N. Sebe and M. Worring, editors, *Proceedings of the 6th ACM International Conference on Image and Video Retrieval (CIVR 2007)*, pages 486–493. ACM, 2007.
7. C. Diou, G. Stephanopoulos, N. Dimitriou, P. Panagiotopoulos, C. Papachristou, A. Delopoulos, H. Rode, T. Tsikrika, A. P. de Vries, D. Schneider, J. Schwenninger, M.-L. Viaud, A. Saulnier, P. Altendorf, B. Schröter, M. Elser, A. Rego, A. Rodriguez, C. Martínez, I. n. Etxaniz, G. Dupont, B. Grilhères, N. Martin, N. Boujemaa, A. Joly, R. Encficiaud, A. Verroust, S. Selmi, and M. Khadhraoui. Vitalas at trecvid-2009. In *Proceedings of the 7th TREC Video Retrieval Evaluation Workshop (TRECVID-2009)*, nov 2009.
8. J.-L. Gauvain, L. Lamel, and G. Adda. The LIMSI broadcast news transcription system. *Speech Communication*, 37(1-2):89–108, 2002.
9. A. G. Hauptmann and M. G. Christel. Successful approaches in the TREC video retrieval evaluations. In Schulzrinne et al. [12], pages 668–675.
10. D. Hiemstra, H. Rode, R. van Os, and J. Flokstra. PFTijah: text search in an XML database system. In *Proceedings of the 2nd International Workshop on Open Source Information Retrieval (OSIR 2006)*, pages 12–17, 2006.
11. L. Hollink, G. Nguyen, D. Koelma, A. Schreiber, and M. Worring. Assessing user behaviour in news video retrieval. In *IEE Proc. Vision, Image and Signal Processing*, pages 911–918, 2005.
12. H. Schulzrinne, N. Dimitrova, M. A. Sasse, S. B. Moon, and R. Lienhart, editors. *Proceedings of the 12th ACM International Conference on Multimedia*. ACM, 2004.
13. C. G. M. Snoek, K. E. A. van de Sande, O. de Rooij, B. Huurnink, J. R. R. Uijlings, M. van Liempt, M. Bugalho, I. Trancoso, F. Yan, M. A. Tahir, K. Mikolajczyk, J. Kittler, M. de Rijke, J.-M. Geusebroek, T. Gevers, M. Worring, D. C. Koelma, and A. W. M. Smeulders. The MediaMill TRECVID 2009 semantic video search engine. In *Proceedings of TRECVID 2009*, 2009.

14. C. G. M. Snoek, K. E. A. van de Sande, O. de Rooij, B. Huurnink, J. C. van Gemert, and J. R. R. e. a. Uijlings. The MediaMill TRECVID 2008 semantic video search engine. In *Proceedings of TRECVID 2008*, 2008.
15. S. Westman, A. Lustila, and P. Oittinen. Search strategies in multimodal image retrieval. In M. Lalmas, A. Tombros, P. Borlund, J. W. Schneider, D. Kelly, J. Feather, and A. P. de Vries, editors, *IiX*, volume 348 of *ACM International Conference Proceeding Series*, pages 13–20. ACM, 2008.
16. P. Wilkins, R. Troncy, M. Halvey, D. Byrne, A. Amin, P. Punitha, A. F. Smeaton, and R. Villa. User variance and its impact on video retrieval benchmarking. In *CIVR '09: Proceeding of the ACM International Conference on Image and Video Retrieval*, pages 1–8, New York, NY, USA, 2009. ACM.