Measuring the Impact of Temporal Context on Video Retrieval

Daragh Byrne, Peter Wilkins, Gareth J.F. Jones, Alan F. Smeaton and Noel E. O'Connor
Centre for Digital Video Processing & Adaptive Information Cluster
Dublin City University, Glasnevin, Dublin 9, Ireland.

{daragh.byrne, peter.wilkins, gareth.jones, alan.smeaton@computing.dcu.ie}

ABSTRACT

In this paper we describe the findings from the K-Space interactive video search experiments in TRECVid 2007, which examined the effects of including temporal context in video retrieval. The traditional approach to presenting video search results is to maximise recall by offering a user as many potentially relevant shots as possible within a limited amount of time. 'Context'-oriented systems opt to allocate a portion of the results presentation space to providing additional contextual cues about the returned results. In video retrieval these cues often include temporal information such as a shot's location within the overall video broadcast and/or its neighbouring shots. We developed two interfaces with identical retrieval functionality in order to measure the effects of such context on user performance. The first system had a 'recall-oriented' interface, where results from a query were presented as a ranked list of shots. The second was 'context-oriented', with results presented as a ranked list of broadcasts. 10 users participated in the experiments, of which 8 were novices and 2 experts. Participants completed a number of retrieval topics using both the recall-oriented and context-oriented systems.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems - *Evaluation/methodology, Video;* H.3.3 [**Information Storage and Retrieval**]: Information search & retrieval - *Information filtering, search process*

General Terms

Design, Experimentation, Human Factors.

Keywords

Video Retrieval, User-evaluation, content-based video retrieval.

1. INTRODUCTION & BACKGROUND

During the presentation of results in modern video search tools, users are normally offered a single representative keyframe from which they must decide if a returned shot is relevant to their information need. Ideally, the selected keyframe conveys the shot's core concepts, but this is often not the case. Providing only a keyframe can make it difficult for a user to judge the relevance of an individual shot, as a single keyframe cannot convey all the temporal activity and objects potentially contained within the video. Supplementary information is often provided in addition to a keyframe to more comprehensively portray results. This additional context information is designed to help users more fully interpret search results and allow them to better distinguish individual shots from one another. A wide range of contexts may be provided for individual video search results. We now explore some of these.

Online video search tools such as YouTube often provide a text-based summary of the contents within the clip. These are user generated, placing the onus on the content creator to provide a good description of the video. This is often not the case and other approaches seek to automatically generate such text-based contextual overviews. Alternative approaches to presenting user-generated text-based descriptions of the clip are to present sources of automatically garnered text-content. Físchlár-News [12] presents text from closed captions, a synchronized audio transcript of television dialogue included with the broadcast transmission. Automatic speech recognition (ASR) has also been used to deliver text based contextual descriptions for video search results [7]. The *Físchlár Digital Video System* [8] presented at TRECVid 2004 uses context-highlighting to draw user attention to sections of the ASR text that match the users query, providing further contextual cues to the relevance of ranked shots.

Another approach to providing context is to offer some form of user assessment within search results. Simple user interface artifacts such as star ratings on YouTube search results [20], allow searchers to assess relevance of a result based on relevance to previous searchers with a similar information need. Normally these modes of context are present in online or community-based retrieval tools as they require large volumes of user feedback for reliability. Where the search task or collection is domain specific, such contextual cues may be difficult or inappropriate to provide.

Display of semantic knowledge within results presentation is also used to augment the keyframe and display additional context metadata associated with the returned clip. By matching the visual features of each frame within the clip to the properties of known concepts (such as indoors, outdoors, people, crowd, etc.) the probability of a given concept's occurrence in the frame can be determined [16]. The IBM Research TRECVID-2006 Video Retrieval System [4] enabled the aggregation of search results into *semantic groups* based on the likely presence of such concepts, determined in this way.

Finally and of increasing popularity is the use of *temporal context* information within results display. This is particularly prevalent in search systems for broadcast-based collections. Within a broadcast collection, each individual broadcast is subdivided into

CIVR'07, July, 2007, Niagra Falls, Ontario, Canada.

a series of smaller units or 'shots'. Whilst a shot may contain a self contained piece of information from within that broadcast, they are often related to their preceding and succeeding shots, which may also contain relevant information. For example, within a news story covering a visit from a foreign dignitary we can perhaps expect to see a shot of that person leaving an airplane, then greeting other political figures and next giving a speech. While these are all distinct shots within the broadcast, they are clearly related and should a user engage in a video search seeking items about that particular dignitary or of political meetings or of foreign state visits, all of these temporally adjacent shots could be of relevance to their information need.

Temporal context in the presentation of results attempts to capitalise on the likely semantic relationship among adjacent shots within a broadcast. By framing the results with their preceding and succeeding shots, the searcher can better understand the concepts being conveyed, the temporal progression within the presented shots and they can often find additional relevant material in the adjacent frames.

Yang and Hauptmann [19] have explored the use of temporal consistency within the ranking of search results. They define this consistency as "*the tendency that the relevant shots ... appear in temporal proximity*" for a given semantic concept or query. They note that while the degree to which relevant items are temporally proximal is dependent on the topic, temporal context is extremely useful in video retrieval. Their investigation was carried out using the TRECVid 2003 video collection [13] which consisted solely of broadcast news footage. They considered temporal adjacency within their study however they did not consider 'temporal neighbourhoods' (we define temporal neighborhoods in Section 4.1).

Several interactive video search systems now include temporal context within results presentation. The MediaMill RotorBrowser [5,17], presents retrieved shots as a visual thread along the horizontal axis. Other axes map out supplementary information or alternative matches to the shot, one of which allowed users to browse the temporal neighbourhood for a selected shot. MediaMill's ForkBrowser [17] and CrossBrowser [15] also incorporate thread-based visualisation of temporal context for shots within a collection. In contrast, Heesch et al [10] present a straightforward approach to the inclusion of temporal context. Within their system they include a fisheye display of temporal neighbours for a selected shot in a panel below the main results. Hauptmann and Christel [9] have surveyed several other TRECVid search systems that make use of temporal context in results.

The number of systems including temporal context within result presentation speaks to assumed benefits within the search process. Given its likely importance within video retrieval we conducted an experimental investigation into the impact of presenting temporal context information on the outcome of interactive video search.

The experiments outlined within this paper were conducted as part of the TRECVid 2007 interactive search task [14,11]. This task requires a set of users to search for a maximum of 15 minutes through a user interface in order to locate video content of relevance to a provided topic. It is important to note that the content of the TRECVid 2007 search collection differs considerably from previous years and therefore from the data used in the evaluation conducted by Yang and Hauptman [19]. While previously broadcast news footage was used, data for the 2007 search tasks comprised approximately 100 hours of news magazine, science news, news reports, documentaries, educational programming, and archival video almost entirely in Dutch [11]. The 2007 collection consequently contains a less regimented format of television content providing a suitable and semantically broader collection to assess impact of temporal context within search result presentation.

In order to address our research question, whether temporal context aids searching within a video archive, we assessed users' performance using a search system which presented results in one of two extremes: a mode with no contextual information but many shots, which was dubbed *recall-oriented,* and a mode with much temporal context, called *context-oriented* which displays results within an entire broadcast and employs an iTunes-esque CoverFlow visualisation. The two systems realise different affordances within interactive search ideally suiting them towards our experimental goals. The evaluation was conducted using 10 participants each of whom completed either 12 or 24 topic-based searches. The experimental setup is outlined in Section 3 following which we discuss the outcomes of our investigation.

2. SEARCH SYSTEM

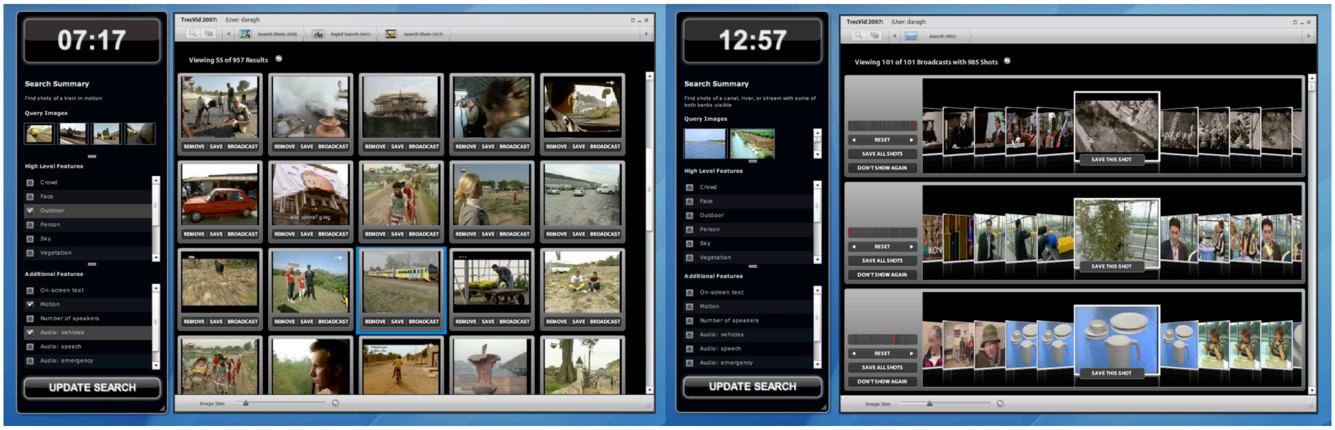
To examine the role of context, K-Space designed two user interfaces, known as the recall-oriented system, and the context-oriented system. Apart from sharing the same retrieval engine, both systems also shared a common query input panel, topic description panel and saved shot area. The major difference was in the presentation of the results from the underlying retrieval engine. The following sections outline the details of the search engine and interface elements.

2.1 Underlying Search Engine

Dublin City University led a TRECVid 2007 submission on behalf of the K-Space consortium, a large European multi-site grouping with an interest in semantic multimedia information management [18]. Video processing in this system began by selecting every second I-Frame from the video (termed KFrames) and for each of these, several low-level feature descriptors were extracted based on the MPEG-7 XM, including colour layout, colour moments, homogeneous texture, edge histogram and scalable colour. K-Frames were also segmented into regions using a Recursive Shortest Spanning Tree (RSST) approach [1], and the same set of MPEG-7 features extracted for each region. Using this set of low-level features, several K-Space partners developed automatic detectors for semantic concepts for each shot. These included for example sports; outdoor; building; mountain; waterscape; waterfront; maps; building; car; desert; road; sky; and snow. Additional detectors included: number of faces visible; camera motion detection; and 17 classes of audio type.. All of these were then combined in the user interface for the system. A more detailed description of the retrieval engine can be found in the K-Space TRECVid paper [18].

2.2 User Interface

For the video retrieval system a rich interactive interface was developed using Adobe Flex [3] and Adobe AIR [2] technologies, allowing for the delivery of a highly interactive desktop search experience. The search interface offers two alternative presentations of search results and both share functionality, navigation and interaction in order to ensure consistency.



07:17

TrecVid 2007: (User: daragh

Viewing 55 of 957 Results

Search Summary

Find shots of a train in motion

Query Images

High Level Features

☐ Crowd
☐ Face
☑ Outdoor
☐ Person
☐ Sky
☐ Vegetation

Additional Features

☐ On-screen text
☑ Motion
☐ Number of speakers
☑ Audio: vehicles
☐ Audio: speech
☐ Audio: emergency

12:57

TrecVid 2007: (User: daragh

Viewing 101 of 101 Broadcasts with 985 Shots

Search Summary

Find shots of a canal, river, or stream with some of both banks visible

Query Images

High Level Features

☐ Crowd
☐ Face
☐ Outdoor
☐ Person
☐ Sky
☐ Vegetation

Additional Features

☐ On-screen text
☐ Motion
☐ Number of speakers
☐ Audio: vehicles
☐ Audio: speech
☐ Audio: emergency

Figure 1. Recall-Oriented vs. Context-Oriented Results Presentation

2.2.1 Common Functionality

2.2.1.1 Rapid querying

A rapid query is a query issued to the search engine using only a single example image. This approach offers the user the ability to uncover shots within the collection that are close matches to a single exemplar image. Such queries can be issued from all areas of the user interface including query images presented in the sidebar, the topic example images on the search screen, saved shots, or any result item. Rapid queries are initiated by double clicking an image and the search results are presented as normal.

2.2.1.2 Sidebar panel

A sidebar panel (see Fig 1) is presented during all operations within the interface to aid the user during their search activities by providing supplementary information on the current search focus and allow the user to refine their search and search results based on their information need. Query reformulation is possible from this screen. As part of the TRECVid benchmark experiments users are limited to a total search time of 15 minutes per topic. The sidebar presents a countdown timer, which during a search counts down to zero after which the search topic is completed and the user is returned to topic selection.

The detection of concepts such as the presence of on-screen text, crowds, faces, sky and buildings have been previously determined within each shot and offer a means by which the results can be filtered. Users can set the filters to positive, negative or off either on the sidebar or at query formulation. For instance, for a query about buildings, a user may set the 'buildings' and 'outdoor' concepts to positive and the face concept to negative. Non-matching elements are made highly transparent but remained in place and visible, allowing the user to traverse the results list with attention drawn to the brighter non-filtered elements.

2.2.1.3 Navigation & Search History Breadcrumbs

At the top of the main application window the users are provided a toolbar with various navigation options (Fig 2). On the far left they are provided with a button to display the search screen (see 2.2.1.4) and another which when selected displays the user's saved shots (see 2.2.1.5). The interface displays 'breadcrumbs' allowing a user to visualise their search path and by simply clicking on an item they can revisit the results of past searches or backtrack if they reach a dead-end trail.

2.2.1.4 Query Formulation

After a topic search begins, users are immediately brought to the search screen (Fig 3). This is divided into two main sections: on the left are basic search options while on the right are advanced options. At the top left is the editable query input text box. It is initially seeded with the TRECVid topic definition text. Beneath this area are visual examples provided for the topic by TRECVid that can be used to seed the query. Clicking on an image will toggle its inclusion. By default all example images are set to on.

On the right side of the screen and as on the sidebar, the user can set the concept filters to be applied to the result set returned from the search. The user, who is more familiar with low-level visual indexes, can modify those used in the visual search. Six indexes are available and are all used by default. Users can also opt to include any available saved shots as visual examples in their query from this screen.

2.2.1.5 Review Saved Results

As a TRECVid search task is to find as many relevant shots as possible, it is an advantage to be able to review shots saved so far, and remove shots not appropriately matching the topic definition.

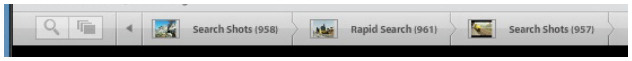


Figure 2. Navigation and search breadcrumbs

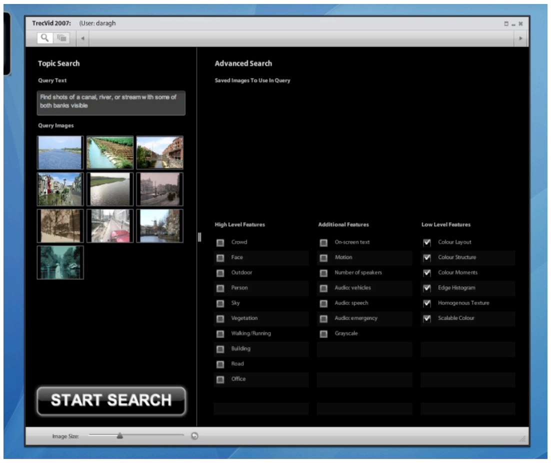
2.2.2 Recall-oriented Results Presentation

The unit of presentation in the recall-oriented interface is an individual shot and in this format users can rapidly iterate through a large set of results. No context for any individual shot is presented. Instead a more traditional layout displaying shots, as ranked by the search engine, is presented. This is considered optimised for recall as the interface attempts to maximise searcher performance by presenting a large number of shots within the available space (Fig 1). If the search engine performs well enough then the searcher should have enough relevant results on screen to iterate through. For each shot a static keyframe is displayed until the user hovers over the keyframe, at which point playback of the shot begins. This is achieved through rapid display of the original broadcast KFrames. No audio is available in shot playback. At the bottom of each shot is a button to save it.

2.2.3 Context-Oriented Results Presentation

The second presentation mode is context-oriented (Fig. 1). Matched shots are returned within a list of ranked broadcasts with each horizontal line in the presentation represents a unique broadcast within the collection. On display of results, each broadcast it is centered on its top-ranked shot. The major benefit of this mode is that it displays high levels of temporal context to the user, showing the preceding and following shots for all matched results.

Like the recall-oriented mode, the context-oriented mode plays back a shot when a user hovers over its keyframe. Ranked results are visually highlighted in the broadcast and users navigate through the broadcast by clicking on a shot of interest which will then centre the broadcast on that shot. Alternatively to the left of the broadcast, a user can click on the arrows to jump through the broadcast five shots at a time or click directly on the timeline to move to a new location. A reset button is provided to re-centre the broadcast on the top ranked shot.



Query Text: Find shots of a canal, river, or stream w ith some of both banks visible

Query Images

High Level Features	Additional Features	Low Level Features
□ Crowd	□ On-screen text	☑ Colour Layout
□ Face	□ Motion	☑ Colour Structure
□ Outdoor	□ Number of speakers	☑ Colour Moments
□ Person	□ Audio: vehicles	☑ Edge Histogram
□ Sky	□ Audio: speech	☑ Homogenous Texture
□ Vegetation	□ Audio: emergency	☑ Scalable Colour
□ Walking/Running	□ Grayscale	
□ Building		
□ Road		
□ Office		

Figure 3. Query Formulation Screen

3. EXPERIMENT OVERVIEW

To examine the role of context, we employed the two presentation modes described previously. Each mode was designed to take an extreme, one with no context and the other presenting context within an entire broadcast, with the only difference between the systems in the presentation of the results.

Ten users participated in the experiment, of which eight were novices and and two were familiar with the search system but were not exposed to the test collection (thereby allowing the two expert retrieval runs to be conducted). It was expected that experts, being more familiar with the systems, should be able to maximise performance compared with novices who would likely still be learning the system during the evaluation. One expert was assigned to the context-oriented results system while the other used the recall-oriented system. Experts were asked to complete one full run of all topics using their assigned system exclusively. While one might argue observed differences may be owing to the experts and not the systems, we believe that the experts used are wholly representative of real world domain experts and provide valueable insights as part of this study. As such their use of the system are considered within the discussion of results.

The remaining eight participants, whilst familiar with content-based retrieval, had not been involved in the TRECVid 2007 activity and as such were unaware of the details of either interfaces or the retrieval system. All users had some previous experience with the use video retrieval systems particularly with online video retrieval systems such as YouTube.

Novice users were asked to complete a total of 12 topics, 6 topics per presentation mode. As novice users had no previous experience with the interface or search system, they were asked to familiarise themselves with the search system for up to 30 minutes before completing 12 topics. For this training, users were provided with two sample topics - one per presentation mode. As the recall-oriented system generally conforms with traditional video retrieval interfaces, it was expected that users would be more quickly and easily able to learn this mode. The context-oriented system was novel, and in order to mitigate against any learning effect which might be noticed in a user's earlier topics a Latin Squares arrangement was used.

For ease, experiments were conducted on the individual participants' desktop computers. Most users within the research group use reasonably high specification machines so system performance across users was consistent and was not a factor within the results. Users were allowed to complete their assigned topics at their convenience over a one week period but were encouraged to do so without interuption. During the experiment and while progressing through their topics, the system actively logged all user activity. Logged activity included: the start and completion of a topic; details of searches performed; the use of concept filters; the saving or removal of shots; keystrokes; mouse-clicks; etc. Additionally a short questionnaire surveying user's experiences with the interfaces was administered.

4. RESULTS AND DISCUSSION

From the ten participants, 4,570 saved shots were recorded across the 24 topics from both systems. On average each participant saved 31.736 shots per topic (30.12 in recall and 27.37 in context, see Table 4). Both systems performed similarly recording approximately the same number of saved shots on average per topic with recall-oriented having 32.36 saved shots and context-oriented having 31.11 (see Table 3). Deeper analysis highlights distinct differences within the search system variants and offers clues as to the importance of temporal context within search results. This is presented in the following subsections,

4.1 Terminology used in discussion of results

Within the following sections there are number of terms used in reporting the experimental findings that should be clarified.

Temporal Sibling: For a given shot deemed to be relevant to a semantic concept or query topic, a temporal sibling exists if either the directly preceding or succeeding shot is also relevant. Temporal siblings are a subset of the set of temporal neighbours.

Temporal Neighbour: For a given shot deemed to be relevant to a semantic concept or query topic, a temporal neighbour exists if either of the preceding or succeeding shots within a given range is also relevant. The context-oriented mode typically presents at least 5 adjacent shots on-screen so for our investigation we consider neighbours to be within 5 adjacent shots. Temporal neighbours may also be referred to as temporal adjacency. For simplicity we consider temporal neighbours only once. For example in a topic with 6 relevant shots, 5 of which are adjacent, we would consider there to be 5 temporal neighbours rather than each relevant shot having 4 neighbours (totaling 20).

Temporal Context Score: Within results exploration, we employ a simple calculation to determine the amount of temporal consistency existing within results for a given topic-based query. This was calculated by dividing the number of temporal neighbours discovered for that topic by the total number of shots judged to be relevant.

4.2 Analysis of Temporal Consistency within the Collection

In order to ascertain the extent to which temporal context may affect search performance and to better interpret the results of the experiment, the presence of temporal consistency within the assessor judged relevant shots was examined. As part of the TRECVid Evaluation Benchmark, all participating sites provide details of the shots they deem to be relevant based on the outcome of their search experiments. From this set of shots, NIST assessors judge each shot's relevance to the search topic. While this relevance ground-truth may not be exhaustive or cover the collection entirely and while it is dependent on the performance of participating systems, it provides a reasonable indication of relevance within the collection.

While in our experiments we examine only *participant-judged* relevance, this review of the 2007 collection from the *assessor-judged* relevance is not without purpose. Our users cannot be expected to locate and judge every possible relevant item within the available time allotted. Additionally it is known that characteristics of individual topics affect the resulting performance of the search (known as *topic effect*) and the amount of temporal consistency present [19]. As such examination of the entire collection's relevance is worthwhile to provide an indication of the topic effect, its difficulty and the degree to which this will impact on participant performance across topics.

The details of this analysis are provided in Table 1 and included the number of assessor judged relevant shots per topic and the number of unique broadcasts those shots exist in (equating to broadcast coverage), in addition to the temporal aspects of those saved shots. Temporal context is on average 0.79 (median 0.84) but despite reasonably high temporal consistency in all topics, it is seen to be highly variable across topics with large changes in the numbers of broadcasts covered, temporal siblings, temporal neighbours and temporal consistency. This supports the previous findings of Yang and Hauptmann [19]. Topic 0218 is shown to exhibit the highest temporal context, also with a relatively low number of broadcasts covered by the relevant set. Thus topic 0218 offers a good example for detailed comparison between the recall and context oriented systems (see Section 4.6) as the benefit of presenting context should be optimal in this case.

Table 1. Temporal Context for Assessor Judged Relevance

Topic No.	Relevant Shots	Broad-casts	Temp. Siblings	Temporal Neighbours	Temporal Context
0197	46	28	10	21	0.46
0198	185	64	57	114	0.62
0199	1150	98	475	966	0.84
0200	105	22	68	89	0.85
0201	195	43	80	148	0.76
0202	49	19	24	34	0.69
0203	51	14	32	44	0.86
0204	174	22	117	156	0.90
0205	108	14	69	95	0.88
0206	330	36	202	295	0.89
0207	257	40	158	217	0.84
0208	74	21	40	56	0.76
0209	327	34	167	278	0.85
0210	18	9	7	12	0.67
0211	15	10	5	8	0.53
0212	77	16	34	62	0.81
0213	389	46	126	322	0.83
0214	255	33	154	227	0.89
0215	145	18	99	129	0.89
0216	57	18	25	43	0.75
0217	112	22	51	95	0.85
0218	374	25	303	353	0.94
0219	6	2	5	5	0.83
0220	205	25	139	176	0.86

As mentioned, the examination of temporal context within assessor judged relevance offers an indication of performance of the systems. This comparison between the average number of saved shots (user judged relevance) for each system against the temporal context score is presented in Table 2. While it is clear that participants within the experiment have only discovered a fraction of the total relevant shots per topic, this is not unexpected given the time constraints imposed and the volume of shots (approximately 20,000) within the collection. Interestingly, instances where the context-oriented system outperformed the recall-oriented system by 10% or more tended to be close to or above the median temporal context for all topics (with the exception of topic 0216), indicating that there is an impact on performance as a result of presenting content when temporal consistency and neighbourhood is high. However, this is not a consistent performance gain and requires deeper analysis as there may be other factors at play, such as topic difficulty or type.

It is worth noting that as the recall-oriented does outperform the context-oriented system in 11 of the 24 topics. Given the volume of results a user can interrogate in the recall-oriented system, one might expect that that it should outperform in all topics. This is shown to be far from the case, hinting towards a performance benefit received from providing temporal context information.

However there is little consistency in the performance of context-oriented relative to the recall-oriented. Topic 0204 shows high temporal consistency (0.897) yet recall-based presentation outperforms context by over 50; topic 0215, with similarly high consistency (0.890) performs in reverse (context oriented system outperforms by over 50%) and the implication is that the topic may have significant impact on the usefulness of temporal context and the benefit gained from presenting temporal context in results may be limited to specific topic types.

4.3 General Observations & Search Strategies

The following sections explore general user behaviour within both results presentation modes. The systems are compared using three general criteria to make determinations as to how participants engaged their searches and exploration of the collection, namely: the number of times filters were applied; the number of searches performed; and the number of saved items.

4.3.1 Use of filters

Semantic concept filters can be toggled within the system to highlight search results matching a user's criteria. It is important to note that within this experiment filters were non-destructive and did not remove results but rather made filtered results appear transparent. This preserved temporal context within the context-oriented system and by fading the non-matching results, users were able to visually interrogate the relevance of displayed shots.

This filtering mechanism offered the same affordances to users within both systems.

While users indicated in feedback that this was counter-intuitive, particularly in the recall-oriented system, filters appear to have been extensively and evenly used within both systems (see Table 3). Interestingly however, filters appear to be used far less by expert users (approximately 15 filter toggles per topic compared to 39 toggles by novice users). The most probable explanation for this is due to a learning effect for novices. Novices are likely to be unfamiliar with the effects of the filters and need to experiment and learn their effect on search results. This makes them more likely to need to refine and correct errors in filter application, accounting for higher use of filters among novices.

Table 2. Temporal Context within the collection compared with system performance as average saved shots per topic. (Instances where the context-oriented system B outperformed System A by 10% or more are highlighted)

Topic	A	B	Context	Topic	A	B	Context
0197	5.3	2.7	0.457	0209	40.0	21.7	0.850
0198	13.0	8.0	0.616	0210	4.0	2.3	0.667
0199	30.3	**49.0**	**0.840**	0211	4.7	3.7	0.533
0200	14.0	**16.3**	**0.848**	0212	41.0	29.0	0.805
0201	48.7	50.0	0.759	0213	61.0	65.0	0.828
0202	9.0	8.0	0.694	0214	84.0	62.0	0.890
0203	11.3	**16.3**	**0.863**	0215	27.0	**42.7**	**0.890**
0204	66.3	30.7	0.897	0216	10.7	**20.7**	**0.754**
0205	14.0	8.7	0.880	0217	26.0	22.3	0.848
0206	46.3	**52.0**	**0.894**	0218	96.0	**119.3**	**0.944**
0207	48.7	51.3	0.844	0219	5.3	**6.3**	**0.833**
0208	22.3	17.0	0.757	0220	47.7	41.7	0.859

Table 3. General Use Statistics for Both Systems

	Avg. Filters	Avg. Searches	Avg. Saves
System A – All	31.03	7.92	32.36
System A – Expert	15.54	11.38	39.08
System A - Novice	38.77	6.19	29.0
System B - All	31.51	3.78	31.11
System B – Expert	14.88	2.38	42.33
System B - Novice	39.83	4.48	25.5

Table 4. General Use Statistics for each participant *(where NE=Non-Expert User, Ex= Expert User)*

User	System A			System B		
	Avg. Filters	Avg. Searches	Avg. Saves	Avg. Filters	Avg. Searches	Avg. Saves
NE 1	32.17	5.67	22.17	43.83	7.17	11.83
NE 2	49.67	14.50	27.67	54.33	4.83	23.83
NE 3	96.67	6.67	16.33	67.00	4.83	36.00
NE 4	11.17	2.83	46.50	21.17	3.83	28.50
NE 5	27.17	4.67	39.67	34.17	4.17	11.67
NE 6	30.33	4.00	24.00	28.83	4.00	40.17
NE 7	21.67	5.00	35.67	33.00	2.17	30.33
NE 8	41.33	6.17	20.00	36.33	4.83	21.67
Ex 1	15.54	11.38	39.08	▨	▨	▨
Ex 2	▨	▨	▨	14.88	2.38	42.33
Mean	36.19	6.76	30.12	37.06	4.25	27.37

Table 5. Use of Rapid Queries across systems.

	System A		**System B**	
	Queries	**Rapid Queries**	**Queries**	**Rapid Queries**
All	7.92	3.25	3.78	1.33
Novice	6.19	2.38	4.48	1.33
Expert	11.38	5	2.38	1.33

4.3.2 Number of searches

During a search topic a participant is expected to explore the collection based on that topic. It is expected that a reasonable number of searches would be performed per topic in order to achieve this. Different approaches to this exploration are highlighted by a marked difference between the number of searches issued across presentation modes and across user types.

Double the number of searches was performed in the recall-oriented presentation mode (7.92 on average per topic) as compared with the context-oriented system (3.78 per topic). This trend is again seen between experts and novices. In System A (recall oriented system), both types of user perform at least double the number of searches (see Table 3). The most marked difference is between the expert users. The recall-oriented expert performing on average 11.38 searches per topic, compared with the 2.38 searches of the context-oriented expert.

This can also be considered in terms of the time spent exploring each result-set. The users were allocated 15 minutes to conduct their search, so for the context expert who performed an average of 2.38 searches, typically slightly over 6 minutes were spent exploring each returned result-set. Due to the volume of context provided, this large exploration time can be expected with users having far more information to review for each matched shot. The recall expert issues a far higher number of queries (11.38 per topic) meaning approximately 75 seconds were spent exploring each result-set. This strongly supports users within the recall-oriented system employing rapid iterative review of the search results while more exhaustive and explorative inspection of broadcasts occurs within the context system. The large exploration time per result-set within the context system additionally suggests that the temporal context provided is heavily, perhaps even exhaustively, used in surveying the collection. Despite not issuing the same number of queries, and likely performing exhaustive exploration of the provided context, it has been shown that the context-oriented system performs well, even outperforming the recall oriented system for certain topics indicating that context provision does not hinder performance.

Rapid queries are issued when a user finds a shot of interest and asks the system to provide a list of results that are similar to it. Within the recall system a large portion of all searches issued are rapid queries (see Table 5) particularly for the expert user. In contrast, within the context system, approximately just one rapid query is issued per search topic. This favouring of rapid queries within the recall system points to exploration of 'information trails' within the result-set.

4.3.3 Number of saved shots

There is no discernable difference or trend in the number of saved shots from novices using either system despite a mild decrease in performance from System A (recall) to B (context). Conversely, performance among experts mildly increases from System B to A. Given the extra amount of context available for each item in results set, and an expected learning effect among novices in

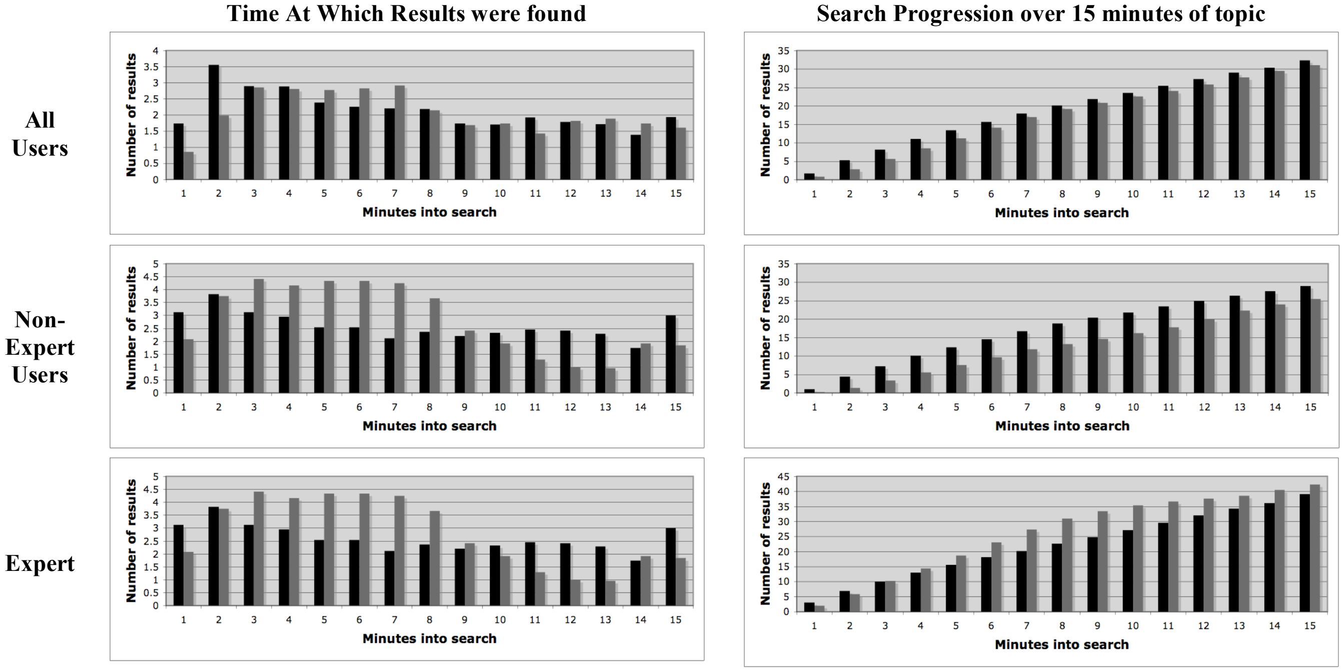


Figure 4. Temporal Aspects of Saved Shots for the Two Presentation Modes
(where recall-oriented system displayed in black and context-oriented displayed in grey)

System B, this is mildly encouraging for the presentation of context information, indicating the presentation of temporal context may be slightly better suited to expert users.

4.3.4 Summary

It was expected in the design of the two presentation modes that System A would support rapid iteration through a result-set leading to high recall while System B in presenting matched results in the context of the entire broadcast would provide extra context to the user requiring a more explorative review of temporal neighbourhoods within the broadcast. The usage data presented seems to support that the systems were used as expected with users spending far longer exploring individual result-sets in System B, as a result of the available context. We assert that they explore in a breath-wise manner across the context of a matched result rather than depth-wise into the set of matched results. While they are not interrogating the results as deeply as their recall counterparts they are achieving comparable performance. This suggests that temporal context is significantly and beneficially augmenting the number of relevant items while minimizing user effort (where effort is search formulation).

4.4 Time to find results

Figure 4 presents a comparison of the time taken for users to discover relevant items (indicated by saving a shot) across systems and across user types. For the first two minutes of a search the recall oriented system locates significantly more results than the context-oriented system. 381 saves were recorded within the first 120 seconds for all topics that used System A, almost double the number recorded for System B (205 saves). This difference is particularly evident in the first two minutes of non-expert searches and also within the first minute of expert searches. It should be noted that a factor within the first minute of a topic is the time taken to formulate and issue the query before results are presented. We can infer that experts are searching much more efficiently as they are accessing more relevant results earlier than non-experts. However as the search proceeds, an advantage appears to be gained by the context system. Within minutes 3 to 7, the context system sustains high levels of saves (mean 204 saves per minute). This high level of activity is particularly pronounced for the expert user but while they see high performance in the first half of a context-oriented topic search, performance declines noticeably within the second half.

The initial lag experienced by the context searchers is likely the time taken to interrogate results to find a highly-temporally consistent broadcast. Once found, the searcher locates matches and discovers a large number of temporal siblings and neighbours, which accounts for the subsequent performance increase.

Differences between the activity of non-expert and expert searchers on the context-oriented system are notable. Given that novices are known to perform more searches than experts on the context system, but yield lower performance in this presentation mode, we can infer that they are not mining temporal consistency to the same degree. Overall though, searches across systems progress similarly over the 15 minutes, though with context-oriented novices underperforming against their recall counterparts and the context-expert out performing the recall.

4.5 Broadcast Coverage & Temporal Context

Our expectation is that users of the context-oriented system will survey results breath-wise using the context data provided. Thus

Table 7. Average Broadcast coverage

	All Users	Experts	Novices
System A	16.083	10.333	12.792
System B	8.375	4.5	6.667

Table 6. Temporal Context For Saved Results

	System A (Recall-Oriented)				System B (Context-Oriented)			
Topic	Broadcasts Covered	Saved Shots	Temporal Siblings	Temporal Neighbours	Broadcasts Covered	Saved Shots	Temporal Siblings	Temporal Neighbours
0197	9	11	0	2	5	8	2	5
0198	18	37	0	25	10	18	2	6
0199	18	88	2	75	10	88	36	78
0200	9	38	4	29	4	27	2	23
0201	27	144	17	117	9	149	23	137
0202	4	22	6	18	4	22	5	18
0203	12	35	2	25	5	47	39	47
0204	27	195	14	168	9	91	44	89
0205	5	38	0	33	3	25	9	22
0206	20	138	4	116	16	142	41	131
0207	27	145	4	122	14	154	66	137
0208	20	66	5	50	11	49	18	38
0209	28	115	2	89	9	65	24	57
0210	5	8	2	5	3	6	4	4
0211	4	6	0	3	4	10	5	8
0212	15	122	11	112	10	86	52	78
0213	28	172	3	137	18	184	39	163
0214	42	251	22	220	18	184	34	168
0215	8	78	2	70	3	126	52	122
0216	18	30	4	15	10	62	19	52
0217	16	75	4	62	10	67	28	59
0218	13	287	0	256	3	350	240	348
0219	2	15	0	15	2	18	12	18
0220	11	108	8	94	11	122	40	110
MEAN	**16.08**	**92.67**	**4.83**	**77.42**	**8.38**	**87.50**	**34.83**	**79.92**

we expect a large number of saved shots in the context system to be returned from the same broadcast. Additionally as it supports exhaustive exploration of context, it is likely that these results will be more homogenous and from fewer broadcasts.

Temporally adjacent shots (siblings and neighbours) are known to be more likely to contain similar content and be visually similar. The total known relevant results, as judged by the NIST assessors, contains large numbers of both temporal siblings and neighbours (see Table 1). Table 6 illustrates the presence of temporally consistent shots within the saved results of both systems. It is not surprising that the recall-oriented system shows a large number of saved shots to be temporal neighbours (an average of 77.42 per topic or approximately 83.5% of all saved shots). Given that the context-oriented system is designed to allow shots of a temporally consistent nature to be easily uncovered it is again not surprising that there is a high number of temporal neighbours saved (an average of 79.92 per topic or 91.5% of all saved shots).

The number of temporal neighbours found in the context-oriented system is higher than in the recall-oriented system and this shows that the provision of context was important in discovering relevant items. Considering that less searches were issued in the context system and on average much more time was spent exploring the result-sets and their associated temporal context (see Section 4.3.2), leads us to conclude that the provision of context is an effective means to find relevant results.

The difference in the number of temporal siblings within the saved results for each system further stresses the importance of providing context information within search results. A modest 4.83 temporal siblings (5% of saved shots; 6.2% of temporal neighbours) were found in the recall-oriented system while 34.83 (39% of saved shots; 43.58% of temporal neighbours) were found on average per topic in the context-oriented system. This shows that searchers within the recall-oriented system miss a large number of relevant temporal siblings. This finding would seem to recommend the inclusion of limited temporal context information (at least temporal siblings) within the presentation of results and their inclusion could offer a significant boost in performance for recall-based systems.

Given the previous indication that users of the context-oriented system were conducting breath-wise exploration through the temporal context, the expectation is that less broadcasts should be traversed but with a higher number of saved shots per broadcast. If the ratio of temporal neighbours to broadcasts covered by the saved results is examined, this position is supported. Table 7 presents the broadcast coverage of the saved shots in order to give an indication as to diversity within the saved shots. Within the recall-oriented system, users explored an average of 16.08 broadcasts per topic while in the context-oriented, half the number of broadcasts were explored (8.38 on average). This means recall-oriented system yields 4.81 temporal neighbours per broadcast while the context-system averages 9.54 neighbours, demonstrating a far higher density of saved results per broadcast within the context system and indicating that the provided matches within a broadcast have been thoroughly explored in terms of their temporal context information.

Experts appear to be far more directed in their exploration of the search results. Table 7 highlights this. On average and for both systems, experts explored fewer broadcasts compared with their novice counterparts. In the recall-oriented system this is attributable to the expert searcher following information trails through 'more-like-these' rapid queries. It is expected that by using rapid queries they would explore more homogenous result-sets and this would attribute lower broadcast coverage and higher numbers of temporal neighbours within saved shots.

4.6 Exploration of Topic 0218

Topic 0218 required users to *"find shots of one or more people playing musical instruments such as drums, guitar, flute, keyboard, piano, etc."* Within the exploration of assessor judged relevance in Section 4.2, it was highlighted as having the most number of temporally consistent relevant shots. This makes it an ideal topic in which to probe more deeply the impact of temporal context within results presentation.

Within Figure 5 an examination of the time at which shots were saved is presented. The general times appear overall consistent with those for all topics (Fig 5). However it diverges significantly from the general timings when examining novices and experts independently. The expert user's saving pattern appears to be fettered with spurts of high activity followed in the second half of the topic search by almost no activity. With novice searchers, we see a long lag until high activity, taking until the 3rd minute for a large number of shots to be saved. This then steadily declines followed by a smaller activity peak towards the end of the search. The highly variable nature of the context-oriented system's saves is in contrast to those of the recall-oriented system, which consistently turns up lower numbers of results throughout the search. Further light can be shed on the behaviour within context-oriented results by examining the broadcast coverage for the saved items (see Tables 8 and 9).

It is clear that broadcast number 157 is extremely important for both systems. This broadcast consists of shots depicting a studio based musical performance, introductions by a presenter and some discussions and interviews throughout. As a result, broadcast 157 contains a high number of relevant shots, of which very large proportions are temporal siblings. It is clear that the provision of context information could be extremely advantageous for this topic and the broadcast coverage and temporal information provided in Table 9 confirms this. From broadcast 157, context-oriented searchers saved 332 temporal neighbours (of which 238 were temporal siblings) of relevance. In contrast only 189 shots were saved from this broadcast by recall searchers and while 172 of these were temporal neighbours, none of them were temporal siblings. This appears unusual given the high density of relevant temporal siblings within this particular broadcast, further highlighting the potential of temporal context information provision to increase searcher performance and effectiveness.

It is further interesting to note that within this topic, the context-oriented searchers saved shots from only 3 broadcasts. Of the 350 saved shots, all bar 2 were temporal neighbours. This showcases the effectiveness of the provision of temporal context in finding relevant results, but not without drawbacks. The provision of temporal context appears to have significantly limited the diversity in the results found. Users of system A, the recall oriented system, save shots from a far broader range of broadcasts, better sampling the collection as a whole and likely resulting in less homogenous results. Conversely as users of the context system are exploring a low number of broadcasts extensively their saved shots are likely to be quite homogenous.

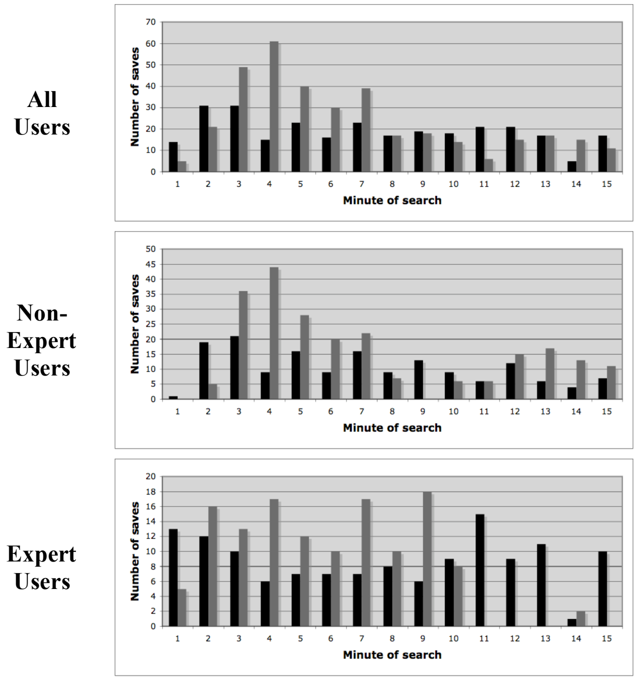


Figure 5. Time taken to save shots within Topic 0218

Table 8. Temporal Context of Saved Items, System A, 0218

ID	Total Shots	Saved Shots	Temp. Siblings	Temporal Neighbours	Temporal Context
157	406	189	0	172	0.91
205	313	2	0	2	1
198	105	33	0	31	0.94
203	329	4	0	4	1
182	92	39	0	34	0.87
128	132	3	0	3	1
124	396	1	0	0	0
195	87	1	0	0	0
149	116	1	0	0	0
204	700	1	0	0	0
217	188	1	0	0	0
172	82	1	0	0	0
163	102	11	0	10	0.91

Table 9. Temporal Context of Saved Items, System B, 0218

ID	Total Shots	Saved Shots	Temp. Siblings	Temporal Neighbours	Temporal Context
157	406	334	238	332	0.99
129	89	2	2	2	1
198	105	14	0	14	1

5. CONCLUSIONS

In this paper we have presented an investigation into the provision of temporal context within search results for users. An analysis of assessor judged relevance within the TRECVid 2007 collection revealed it to be highly temporally consistent, containing a high number of temporally adjacent relevant shots. Thus it is not surprising that a traditional recall-oriented result presentation and a novel presentation of each matched shot's temporal context within its entire broadcast, Both result in users locating large number of temporal neighbours in their saved results. While users are likely find temporal neighbours regardless of provision of context information, there is a distinct difference in the location of temporal siblings (those shots of relevance which are directly adjacent to one another.) The context-oriented system significantly outperforms here. We have also noted differences in the search strategies adopted as a result of providing context to users. Users employ a detailed exploration of the provided context to locate relevant shots while those using a recall-oriented system explore more deeply into the returned results, following information trails and regularly reformulating their queries. Users provided with temporal context in addition spent longer exploring each individual set of results.

The provision of context shows a mild lag in initial performance, likely due to the time taken to locate broadcasts of high temporal consistency and relevance but offer good or high performance from approximately the 3rd to 7th minute inclusive. Overall though, both systems perform equally well. Finally we have demonstrated that the type of user (novice or expert) will impact on the performance of the search and likely the search strategy employed with each system.

ACKNOWLEDGMENTS

This work is supported by the European Commission under contract FP6-027026 (K-Space), the Irish Research Council for Science Engineering and Technology and by Science Foundation Ireland under grant 03/IN.3/I361. We also extend our thanks to the participants in our experiments.

REFERENCES

[1] Adamek T. and O'Connor N. Using Dempster-Shafer theory to fuse multiple information sources in region-based segmentation. In ICIP 2007 - Proceedings of the 14th IEEE International Conference on Image Processing, 2007.

[2] Adobe AIR, http://labs.adobe.com/technologies/air/

[3] Adobe Flex, www.adobe.com/flex

[4] Campbell M., S. Ebadollahi, M. Naphade, A. P. Natsev, J. R. Smith, J. Tesic, L. Xie, and A. Haubold. IBM Research TRECVID-2006 Video Retrieval System. In Proc. TRECVID, Gaithersburg, Md., Nov. 2006.

[5] Cooke E, Ferguson P, Gaughan G, Gurrin C, Jones G, Le Borgne H, Lee H, Marlow S, Mc Donald K, McHugh M, Murphy N, O'Connor N, O'Hare N, Rothwell S, Smeaton A.F and Wilkins P. TRECVID 2004 Experiments in Dublin City University. In Proceedings of TRECVID 2004, Gaithersburg, Maryland, 15-16 Nov. 2004.

[6] de Rooij O, Snoek C.G.M., and Worring M. Mediamill: Semantic video browsing using the RotorBrowser. In Proc. ACM Intnl. Conf. on Image and Video Retrieval (CIVR), pages 649-649. Amsterdam, The Netherlands, July 2007.

[7] Gaughan G, Smeaton A.F, Gurrin C, Lee H and McDonald K. Design, Implementation and Testing of an Interactive Video Retrieval System. MIR 2003 - 5th ACM SIGMM International Workshop on Multimedia Information Retrieval, Berkeley, CA, 7 November 2003. pp 23-30

[8] Gurrin C, Lee H and Smeaton A.F. Físchlár @ TRECVID2003: System Description. ACM Int. Conf. on Multimedia 2004, New York, NY, Oct 2004. pp. 938-939

[9] Hauptmann, A. G. and Christel, M. G. 2004. Successful approaches in the TREC video retrieval evaluations. In Proc. 12th Annual ACM Int. Conf. on Multimedia (New York, NY, USA, October 10 - 16, 2004). ACM Press, 668-675.

[10] Heesch D., Howarth P., Magalhães J., May A., Pickering M., Yavlinsky A., and Rüger S., Video retrieval using search and browsing, In Proceedings of TRECVID 2004, Gaithersburg, Md., Nov. 2004.

[11] Over P, Awad G, Kraaij W and Smeaton A.F. TRECVID 2007 Overview. In Proceedings of TRECVID 2007, Gaithersburg, Md., 5-6 Nov. 2007.

[12] Smeaton A.F, Gurrin C, Lee H, Mc Donald K, Murphy N, O'Connor N, O'Sullivan D, Smyth B and Wilson D. The Físchlár-News-Stories System: Personalised Access to an Archive of TV News. In Proceedings of RIAO 2004, Avignon, France, 26-28 April 2004. pp. 3-17

[13] Smeaton, A.F. and Over, P. TRECVid: Benchmarking the effectiveness of information retrieval tasks on digital video. In Proc. Int. Conf. Image & Video Retrieval, (CIVR), 2003.

[14] Smeaton A.F, Over P and Kraaij W. Evaluation Campaigns and TRECVid. MIR 2006 - 8th ACM SIGMM International Workshop on Multimedia Information Retrieval, Santa Barbara, CA, 26-27 Oct. 2006.

[15] Snoek C.G.M., Everts I., van Gemert J.C., Geusebroek J.M., Huurnink B., Koelma D.C., van Liempt M., de Rooij O., van de Sande K.E.A., Smeulders A.W.M., Uijlings J.R.R. and Worring M. The MediaMill TRECVID 2007 Semantic Video Search Engine. In Proceedings of TRECVID 2007, Gaithersburg, Md., Nov. 2007.

[16] Snoek, C.G.M., Huurnink, B., Hollink, L., de Rijke, M., Schreiber, G. and Worring, M.. Adding semantics to detectors for video retrieval. IEEE Transactions on Multimedia, 9(5):975-986, August 2007.

[17] Snoek C.G.M., van Gemert J.C., Gevers Th., Huurnink B., Koelma D.C., Van Liempt M., De Rooij O., van de Sande K.E.A., Seinstra F.J., Smeulders A.W.M., Thean A.H.C., Veenman C.J., Worring M. The MediaMill TRECVID 2006 Semantic Video Search Engine. In Proceedings of TRECVID 2006, Gaithersburg, Md., November 2006.

[18] Wilkins P. et al. K-Space at TRECVid 2007. In Proceedings of TRECVID 2007, Gaithersburg, Md., 5-6 Nov. 2007.

[19] Yang, J. and Hauptmann, A. G. 2006. Exploring temporal consistency for video analysis and retrieval. In Proc. 8th Int. Workshop on Multimedia Information Retrieval (Santa Barbara, California, USA, Oct. 2006). MIR '06. ACM, 33-42.