International Journal of Imaging Systems and Technology

Content-Based Video Retrieval: Three Example Systems from TRECVid

Journal:	*International Journal of Imaging Systems and Technology*
Manuscript ID:	IMA-07-119.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Smeaton, Alan; Dublin City University, Centre for Digital Video Processing Wilkins, Peter; Dublin City University, Centre for Digital Video Processing Worring, Marcel; University of Amsterdam, Intelligent Systems Lab de Rooij, Ork; University of Amsterdam, Intelligent Systems Lab Chua, Tat-Seng; National University of Singapore, School of Computing Luan, Huanbo; Chinese Academy of Sciences, 4Institute of Computing Technology
Keywords:	video retrieval

Content-Based Video Retrieval: Three Example Systems from TRECVid

Alan F. Smeaton[1], Peter Wilkins[1], Marcel Worring[2], Ork de Rooij[2], Tat-Seng Chua[3] and Huanbo Luan[4]

[1]Centre for Digital Video Processing and Adaptive Information Cluster, Dublin City University, Glasnevin, Dublin 9, Ireland.

[2]Intelligent Systems Lab Amsterdam, University of Amsterdam, Kruislaan 403, 1098 Amsterdam, The Netherlands.

[3]School of Computing, National University of Singapore.

[4]Institute of Computing Technology, Chinese Academy of Sciences, China.

Abstract

The growth in availabile online video material over the internet is generally combined with user-assigned tags or content description, which is the mechanism by which we then access such video. However, user-assigned tags have limitations for retrieval and often we want access where the content of the video itself is directly matched against a user's query rather than against some manually assigned surrogate tag. Content-based video retrieval techniques are not yet scalable enough to allow interactive searching on internet-scale, but the techniques are proving robust and effective for smaller collections. In this paper we show 3 exemplar systems which demonstrate the state of the art in interactive, content-based retrieval of video shots, and these three are just three of the more than 20 systems developed for the 2007 iteration of the annual TRECVid benchmarking activity. The contribution of our paper is to show that retrieving from video using content-based methods is now viable, that it works, and that there are many systems which now do this, such as the three outlined herein. These systems, and others can provide effective search on hundreds of hours of video content and are samples of the kind of content-based search functionality we can expect to see on larger video archives when issues of scale are addressed.

1. Video Search as a MMIR Application

Of all the media to which we now have relatively easy access, video, in digital form, is the one which has the steepest growth curve. Digital TV and set-top boxes, personal video recorders, DVDs and more recently internet video such as through YouTube or FabChannel, all contribute to placing enormous video archives at our disposal, if only we could navigate them effectively. Most of the technical issues associated with the video lifecycle are now solved to all practical intents and purposes. We can easily capture and store video, we can compress it, transmit it, and we can easily render it on fixed or mobile platforms. What remains our greatest technical challenge is being able to navigate it, to

be able to browse it and search it in order to find clips which are of interest or of value to us.

The dominant approach to navigating digital video in large-scale practical applications is to use video metadata, either automatically determined or manually assigned. Automatic metadata includes date and time, and provides limited usefulness when the archives are large. Typically, video is annotated with descriptions which may include title, actor(s), storyline, perhaps even a dialogue script. Publicly available systems such as Open Video[1] or the Internet Movie Database[2] are examples of systems using such metadata only, and which have been in widespread use for large closed archives for some time.

Because we can now easily capture and store video and upload it to the internet for sharing we have many systems where the description of shared video is augmented by user-assigned terms or tags. The Internet Movie Archive[3] and the popular YouTube system[4] are examples of systems where content description is determined mostly by end users directly. This can take the form of user-assigned tags or keywords, or can be user reviews of the video, and all of these are used to provide video retrieval.

While content description from user annotation offers useful navigation possibilities, it is still one step removed from being able to search actual video content directly. Effective use of user annotation relies on manual effort and consistent annotation, and this is not scalable for developing good quality content-based access to large quantities of video. In this paper we concentrate on the application of direct content access to video where user queries are matched directly against video content. We present three example systems which demonstrate differing approaches to content-based video search, each developed in the context of a large scale worldwide benchmarking activity where dozens of video indexing and retrieval systems are benchmarked on the same video dataset.

The rest of this paper is organised as follows. In the next section we summarise the benefits of a common evaluation carried our across a number of research groups and in particular, the annual TRECVid activity. This section is included as background and is followed by an overview of each of three representative video retrieval systems developed by Dublin City University/K-Space, by the University of Amsterdam/MediaMill, and by the National University of Singapore, respectively. These three systems are chosen from over 20 systems for interactive video search developed for the 2007 TRECVid search task. Each have different approaches to the task of content retrieval from video. The similarities and differences between these systems, as well as a report on their respective performance in the TRECVid evaluation, are presented and discussed in section 4, followed by some overall conclusions.

[1] www.open-video.org
[2] www.imdb.org
[3] www.archive.org
[4] www.youtube.com

2. TRECVid: A Benchmarking Evaluation Campaign for Video Retrieval

Evaluation and common benchmarking is important in many kinds of image and vision processing. The development of video compression algorithms, for example, has always taken place in the context of shared and common datasets on which compression proposals can be compared directly. Currently there are several example evaluations for content-based tasks on video including ETISEO (Evaluation du Traitement et de l'Interprétation de Séquences Vidéo)[5] which targeted vision techniques for video surveillance applications involving pedestrians and/or vehicles, PETS (Performance Evaluation of Tracking & Surveillance) (Lazarevic-McManus *et al.*, 2006) which targets object detection and tracking for multi-view/multi-camera surveillance and ARGOS (Joly *et al.*, 2007) which targeted shot boundary detection, camera motion detection, person identification, video OCR and story boundary detection on broadcast TV news, scientific documentaries and surveillance video.

In terms of video retrieval the largest collaborative benchmarking activity for content-based activities is the series of TRECVid workshops, running annually since 2001 (Smeaton *et al.*, 2006). This has involved worldwide participation with over 50 research teams taking part each year in a variety of content-based "tasks" including shot boundary detection, concept or semantic feature detection, automatic summarization as well as content-based video retrieval. In 2007 the data used consisted of educational, cultural, youth-oriented programming, news magazines, historical footage video taken from the Dutch Sound and Vision archive and primarily in Dutch (Over *et al*., 2007*)*. This data had a great variety of subject matter. The volume of video data used varied each year, with 160 hours of MPEG-1 video used in 2006 for example.

The interactive search task involved applying whatever video analysis and indexing tools each participant had to the search data and building their search system around that data. Participants were also able to take advantage of a variety of metadata donations made by the research community to the task and these included (for the 2007 TRECVid cycle alone) a master shot segmentation formatted as MPEG-7, automatic speech recognition output and translation of that into English, low-level features derived from each shot, outputs from 374 semantic feature detectors applied to 2007 data and trained on 2005 data from Columbia University, applied to 2007 data and trained on 2006 test data from City University of Hong Kong, and two sets of manual annotations for 36 semantic features as the result of large-scale collaborative annotation activities (Quénot, 2007), (Jiang *et al.*, 2007).

The definition of the search task required each participating group to submit the results of running each of 24 topics or statements of information need against the search data. Each of the text description of topics is augmented by several illustrative images and/or video clips as exemplars of the information need, corresponding to the scenario where the searcher already has some images/video clips which are relevant to the information need.

[5] www.silogic.fr/etiseo

The shot lists returned by each participant for each topic were pooled together to some depth, duplicates were removed and shots were manually assessed for relevance by the TRECVid organizers. Once this ground truth of relevant shots for each topic was determined, the organizers were then able to compute the absolute performance figures for the submitted runs in terms of precision and recall as measured against the manually assessed pooled ground truth.

In the interactive search variant, participants were allowed to submit a number of runs (up to 6 in 2007) where each topic in each run was limited to the shots deemed to be relevant and found by one person using the participating site's search tool, as found within a 15 minute limit. This simulated the scenario where a searcher has a limited timeframe to find as many shots as he/she can where each shot is relevant to a fixed, unwavering information need. Such a scenario would regularly occur in a newsroom for example, where a production assistant seeks to locate video footage on a news topic to present to a news editor for possible inclusion in a broadcast.

The systems described in the next section of this paper are from three of the twenty-four research groups who participated in the interactive search task in TRECVid 2007 and we describe each system in turn. The three systems were chosen for their variety rather than their absolute performance characteristics in order to illustrate the capabilities of contemporary content-based video retrieval systems.

3. Three Sample Video IR Systems

Each participant in the TRECVid search task normally addresses some research question or issue which is of interest to them, and may run more than one variant of their system in order to submit a number of "runs" which are each assessed manually by the TRECVid organisers. For each run we can compute retrieval performance figures like precision and recall and can average these across the set of topics to give an indicative score of the performance of the system behind each "run". We now describe three systems to illustrate the capabilities of content-based video retrieval within TRECVid.

3.1 Dublin City Universty/K-Space Interactive Video Retrieval

The team from 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 (Wilkins *et al*., 2007). Video processing in this system used every second I-Frame, terming a *K-Frame*, and extracting several low-level feature descriptors 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 (Adamek and O'Connor, 2007), and the same set of MPEG-7 features extracted for each region. Several K-Space participants developed several automatic detectors for semantic

concepts for each K-frame, including *sports, outdoor, building, mountain, waterscape/waterfront, maps*, *face detection*, *17 classes of audio type*, *building, car, waterscape-waterfront*, *desert, road, sky, snow, vegetation, explosion/fire, mountain*, *camera motion, number of* faces visible, *weather, US-flag, boat/ship* and *vegetation*. These were then combined in the user interface for the system.

The DCU/K-Space experiment under investigation in TRECVid was to examine the role of context in the user interface, where context can be described as showing temporally adjacent shots. To examine the usefulness of such context, DCU/K-Space designed two user interfaces, the `shot based' system, and the `broadcast based' system. Both systems, apart from sharing the same retrieval engine, 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 `broadcast based' system takes the idea of context to its maximum by ranking not just individual video shots, but entire TV broadcasts. This presents an alternative to a shot-only presentation of results and allows a searcher to explore the temporal neighborhood of shots. In Figure 1 we see a horizontal line of shots in rows across the results area. Each row is an entire broadcast, with the best-matching broadcast being the first row. When a user issues a query, the ranked list of broadcasts is presented, and within each broadcast, the row will be centered on the highest matching shot within that broadcast. The coverflow-like interface allows for rapid browsing of shots within a broadcast.

Figure 1 shows the user's multimodal query and includes the text "Find shots of a canal, river or stream with some of both banks visible" which is matched against the translation of the automatic speech recognition. Also included are two sample query images which have either been found by the searcher, or form part of the topic definition, and a subset of the available semantic features, in this case *outdoor*. Query images are matched against the K-frames from each shot using the same low-level features mentioned earlier and each of the modalities (text search and image matching) generates a separate ranking of shots. Using a variation on a query-time weight generation techniques (Wilkins and Smeaton, 2006), the independent result lists are merged at query time with weights being assigned to each retrieval expert which approximate that expert's likelihood of providing the most relevant responses to the query. The semantic concepts are then used as filters by the user after a content-based query has been issued and these filters can be set to `positive', `negative' or `off'. In Figure 2 we can see that the user's query has moved on and s/he has found a total of 6 query images but has disabled the semantic concept feature filtering of *outdoor*.

3.2 University of Amsterdam/MediaMill ForkBrowser

The MediaMill team at the University of Amsterdam departed from the traditional cycle of query-browse-query by providing users with video browsers that allow to visualize the entire data set in multiple dimensions, thus facilitating interactive exploration. For TRECVid 2007, the focus was specifically on consolidation of proven interface

components from previous TRECVid editions into a novel browsing environment (Snoek *et al.*, 2007)

The notion of threads was introduced in the ForkBrowser in order to browse through a video data set in multiple directions. A thread is a linked sequence of shots in a specified order, based upon an aspect of their content (de Rooij *et al.*, 2007) including *static threads* which are pre-computed, and *dynamic threads* which are generated on demand. The content of a thread is based on a form of similarity between shots in the data set.

The combination of a time thread with any other thread resulted in the CrossBrowser which proved effective for the TRECVid interactive search tasks in 2004 and 2005 when a single thread was sufficient for the user to find shots which satisfy a topic or information need (Snoek *et al.*, 2007), (Snoek *et al.*, 2006). For topics that require a combination of threads, the RotorBrowser was introduced in 2006 (de Rooij *et al.*, 2007), (Snoek *et al.*, 2006). This allows a user to integrate query results with time, visual similarity, semantic similarity and various other shot-based similarity metrics. While effective, this visualization proved overwhelming for non-expert users. To leverage the benefits of having multiple query methods while simultaneously allowing the user to maintain an overview of their results, an interface was introduced in TRECVid 2007 which combines query by keyword, query by example, query using 572 semantic concepts, query by time and by program, all combined into a framework which is called the ForkBrowser.

The ForkBrowser visualizes results by displaying keyframes based on the shape of a fork. The contents of the tines of the fork depend on the shot at the top of the stem. The center tine shows unseen query results, the leftmost and rightmost tines show the time thread, and the two tines in between show user-assignable threads. For the TRECVid 2007 benchmark two variants of visual similarity threads are displayed. The stem of the fork displays the history thread. Every displayed key frame is taken from a single video shot, and the video shot can also be played on demand by rapidly displaying up to 16 frames in sequence from the originating shot. This helps in rapidly answering queries containing explicit reference to motion or to events. Figure 3 depicts the ForkBrowser while searching for “boats moving past”. The horizontal tine shows shots from the time thread of the program “Klokhuis”, the diagonal directions depict two visual threads to provide the user with similar shots from waterscapes which s/he can browse.

3.3 NUS-ICT/VisionGo

VisionGo is an interactive video retrieval system developed jointly by the National University of Singapore (NUS) and the Institute of Computing Technology, Chinese Academy of Sciences (ICT). The system is designed to maximize the effectiveness of human annotators through the use of an intuitive User Interface (UI), options for multiple feedback strategies and motion icons. In performing an interactive search, the system first uses results from an automated search based on the user’s multimodal query, followed by multimodal fusion to retrieve a ranked list of shots. The fusion uses a combination of text

derived from ASR (automatic speech recognition), high-level features (HLFs) automatically detected in shots, and a combination of low-level visual features and motion (Chua *et al.*, 2007). The user then makes use of an intuitive retrieval interface with a variety of relevance feedback options to refine their search results. In addition, motion-icons are introduced which allow users to see a dynamic series of keyframes instead of a single keyframe during relevance assessment.

To maximize the user's interaction efforts, the intuitive UI is designed for fast keystroke actions with quick previews of previous and subsequent sets of shots in the ranked list of shots. A sample interactive UI is shown in Figure 4. The UI is inspired by high throughput interactive game interfaces, which are mainly keystroke based. The UI displays three images at a time in a central active row, with the previous and next rows in view. Each image corresponds to a single retrieved shot, without any *context* such as previous or following shots.. The user will determine the images' relevance to the query and annotate the positive ones by hitting pre-a defined set of keys on the keyboard. The system captures the user's input and automatically refreshes itself to display the next row of new keyframes in the ranked list. In experiments at the National University of Singapore, the UI enabled a normal user to annotate up to 3,500 shots based on motion icons or 5,000 shots based on static icons, in only 15 minutes.

To allow for more flexibility and to provide a range of options for users to click during relevance feedback, interactive feedback is segregated into three distinct types, namely recall-driven, precision-driven and temporal locality-driven feedback. Each strategy aims at leveraging different aspects of user feedback data. At any time, if the user feels that the search and feedback process is not progressing well, he/she is able to select any other feedback strategy to enhance search performance.

Recall-driven feedback employs general features such as the ASR text tokens and HLFs from relevant shots to perform query expansion. This option has been found to be the most effective in finding many new relevant shots in the initial stage of a search. Given the set of positively annotated shots, this process makes use of text and HLF scores to iteratively adjust the retrieval function. Precision-driven feedback uses motion, visual and audio features in an SVM-based active learning environment targeting at improving precision. It uses active learning to provide long term improvements to classifiers. Fused with a performance-based adaptive sampling strategy, this process continuously re-ranks instances as the user annotates shots as relevant or non-relevant. Finally, temporal locality-driven feedback essentially returns shots from neighboring shots from the positively labeled set, as it is found that positive shots tend to cluster near each other within the same story. Based on these multiple feedback strategies, a user is able to choose the type of feedback that is more suitable based on his/her intuition or experience, in order to maximize performance.

Many visual-oriented queries tend to be associated with objects in motion in the video. It is therefore necessary to provide some information on motion within each shot. Specifically, we construct a summarized clip comprising a sequence of progressive keyframes which can show moving picture information. We call this a motion icon or

micon. Through the use of micons in previewing shots, the user has a clearer idea of what motion information is in the shot and can identify relevant shots more quickly and with more confidence.

4. A Comparison of Performance of the Three Search Systems

Since one of the purposes of the TRECVid benchmarking activity is to compare the performance of various content-based video retrieval approaches, in this section we summarise how the three systems introduced earlier performed in their official TRECVid submissions. It is important to remember that each of the TRECVid participating groups set out to examine some research question and achieved this by comparing performance from among their allowed six runs.

There are many topical research areas within the broad area of content-based video retrieval including (semantic) concept detection, query formulation, algorithms for image/keyframe matching, fusion of individual retrievals (colour, texture, edge based, text based, concept based, etc.), browsing interfaces and the issue of how to effectively incorporate relevance feedback into the user experience. In fact successfully exploiting relevance feedback into a retrieval interface and presenting it as an integral part of the user experience is a major challenge in content-based video retrieval and an issue that the NUS system concentrated upon as its research question. Relevance feedback is a hugely important aspect of the user interaction in video retrieval because video is so rich in terms of content, and any video search system will generally require much interaction with its enduser, through relevance feedback and video browsing, in order for the user to locate relevant shots and so it is an area of much research activity as can be seen in the overview in (Zhou and Huang, 2002) and more recent work, for example in (Tao *et al.*, 2008).

Within the set of up to six runs allowed from each site participating in TRECVid, each group can control the one variable, the endusers, which cannot be controlled in performance comparisons across sites. What this means is that it is acceptable to compare precision and recall figures within a site's runs, but comparisons across sites will have an uncontrolled variable as users will vary in their levels of expertise, experience, or even in their level of motivation for performing the searches.

Notwithstanding the above caveat, there is value for us here in examining the performance of the best runs from each site but only in looking at the absolute values. Figure 7 shows the precision-recall figures for these runs and shows that the three systems are quite comparable – the codes in the legend refer to the official TRECVid run names used to distinguish participants and their submitted runs. What is most noteworthy from the point of view of this paper is the performance of the three systems at the high precision end of the scale, corresponding to the 'early' parts of the user's search. Here we can see that performances for all of the systems are very effective, meaning that each of these systems, and many of the others developed in TRECVid, provide effective tools for helping users locate relevant shots, and these are all based on content-based searching.

5. Content-Based Video Retrieval Conclusions

The three interactive video search systems presented in this paper are both similar and different to each other. The similarities are that each supports a multimodal query from a user – a combination of text, sample images(s) and semantic features – which is implemented by running multiple shot ranking algorithms for each of the modalities and then fusing their outputs together at search time. Each supports a preview of a whole shot by presenting sets of keyframes, called *micons* by NUS and K-Frames by K-Space, to allow a user to determine whether an event of some kind occurs within a shot.

Yet despite these similarities there are huge differences in the interfaces and user experiences among the three systems which have afforded each of them to explore some aspect of the retrieval interaction as an experiment. DCU/K-Space experimented with the effects of local context and within-broadcast impact on retrieval quality; University of Amsterdam/MediaMill experimented with the effects of different threads including a history thread, while National University of Singapore experimented with the effects of different relevance feedback algorithms. Collectively, however, what the three systems demonstrate is that there are now many systems which can provide effective content-based retrieval of video shots from archives of several hundreds of hours of video content, in a fast, effective and user-friendly manner. TRECVid search systems represent the state-of-the-art in content-based video searching yet this is not mainstream in terms of usage by a large population of users. The techniques needed to realize a widespread deployment of this, such as an internet-scale deployment, are under development and represent one of the largest challenges in this field.

Acknowledgements

AS and PW wish to acknowledge support from Science Foundation Ireland under grant 03/IN.3/I361.

References

T. Adamek and N. O'Connor. 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.

TS Chua *et al.* TRECVID 2007 Search Tasks by NUS-ICT. In Proceedings of TRECVID 2007, Gaithersburg, Md., November 2007.

Yu-Gang Jiang, Chong-Wah Ngo, Jun Yang. Towards Optimal Bag-of-Features for Object Categorization and Semantic Video Retrieval. ACM International Conference on Image and Video Retrieval (CIVR'07), Amsterdam, The Netherlands, 2007.

P. Joly, J. Benois-Pineau, E. Kijak, and G Quénot. The Argos Campaign: Evaluation of Video Analysis Tools. In Proceedings of the International Workshop on Content-Based Multimedia Indexing, 2007. CBMI'07, pages 130-137, 2007.

N. Lazarevic-McManus, J. Renno, J. and G.A. Jones, G. A. 2006. Performance evaluation in visual surveillance using the F-measure. In Proceedings of the 4th ACM international Workshop on Video Surveillance and Sensor Networks (Santa Barbara, California, USA, October 27 - 27, 2006). VSSN '06, 45-52.

P. Over, G. Awad, W. Kraaij and A.F. Smeaton. TRECVID 2007 - An Introduction. In Proceedings of TRECVID 2007, Gaithersburg, Md., November 2007.

G.M. Quénot. Active learning for multimedia. In Proceedings of the 15th international Conference on Multimedia (Augsburg, Germany, September 25 - 29, 2007). MULTIMEDIA '07. ACM Press.

O. de Rooij, C.G. Snoek and M. Worring. Query on demand video browsing. In Proceedings of the 15th international Conference on Multimedia (Augsburg, Germany, September 25 - 29, 2007). MULTIMEDIA '07. ACM Press, 811-814.

A.F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and TRECVid. In Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval (Santa Barbara, California, USA, October 26 - 27, 2006). MIR '06. ACM Press, 321-330.

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

C.G.M. Snoek, M. Worring, D.C. Koelma and AWM Smeulders. A Learned Lexicon-Driven Paradigm for Interactive Video Retrieval. IEEE Transactions on Multimedia, 9(2), pages 280-292, 2007.

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

D Tao, X Tang and X Li. Which Components are Important for Interactive Image Searching ? IEEE Transactions on Circuits and Systems for Video Technology, 18(1), pages 3-11, 2008.

P. Wilkins, P. Ferguson, and A. F. Smeaton. Using score distributions for querytime fusion in multimedia retrieval. In Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval (Santa Barbara, California, USA, October 26 - 27, 2006). MIR '06. ACM Press, 51-60.

P. Wilkins *et al.* K-Space at TRECVid 2007. In Proceedings of TRECVID 2007, Gaithersburg, Md., November 2007.

X. Zhou and T.S. Huang. Relevance Feedback in Content-Based Image Retrieval: Some Recent Advances. *Information Sciences Applications*, 148(1-4), 2002, 129-137.

Figure 1: User interface for Dublin City University K-Space Search System

Figure 2: User interface for Dublin City University / K-Space Search System

Figure 3: User interface for University of Amsterdam's Search System

Figure 4: User interface for National University of Singapore's VisionGo Search System

Figure 5: A sequence of multiple keyframes for shot213_62

Figure 6: A sequence of multiple keyframes for shot149_62

Figure 7: Performance figures for three systems from TRECVid

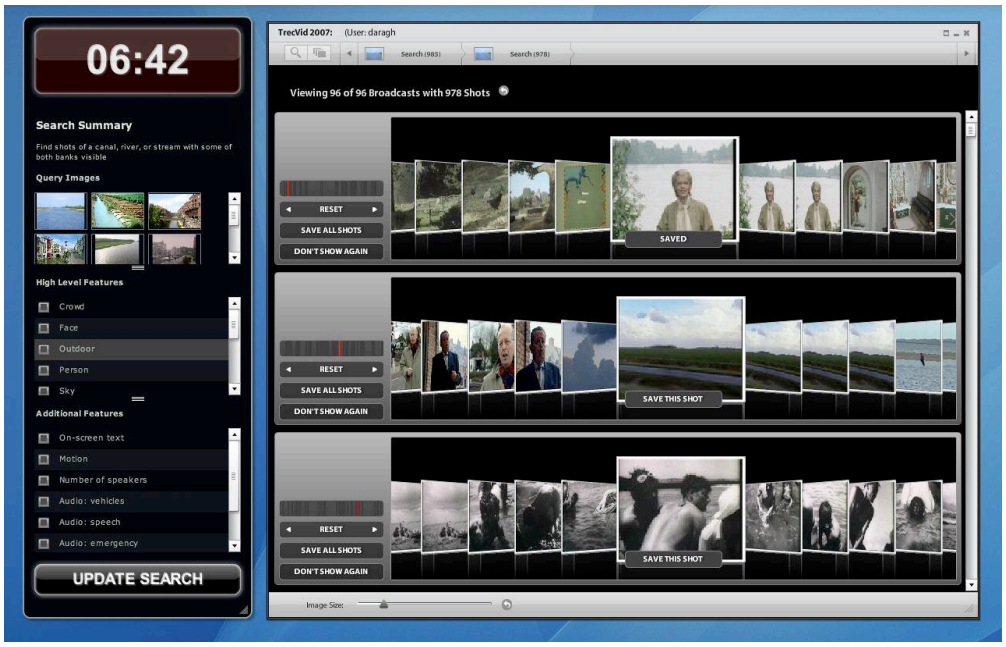


Figure 1: User interface for Dublin City University K-Space Search System
451x287mm (72 x 72 DPI)

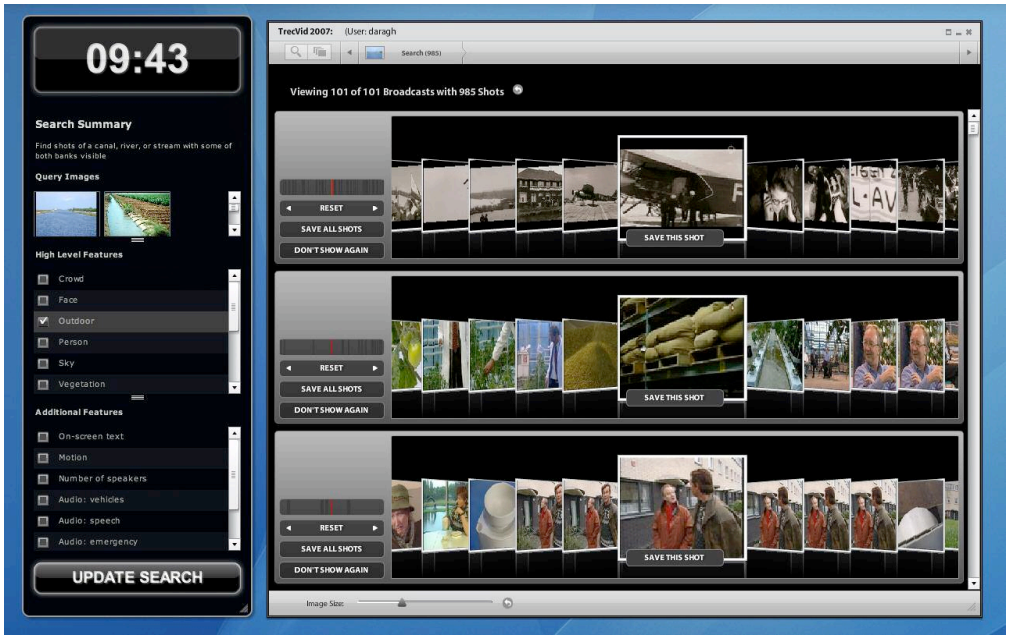


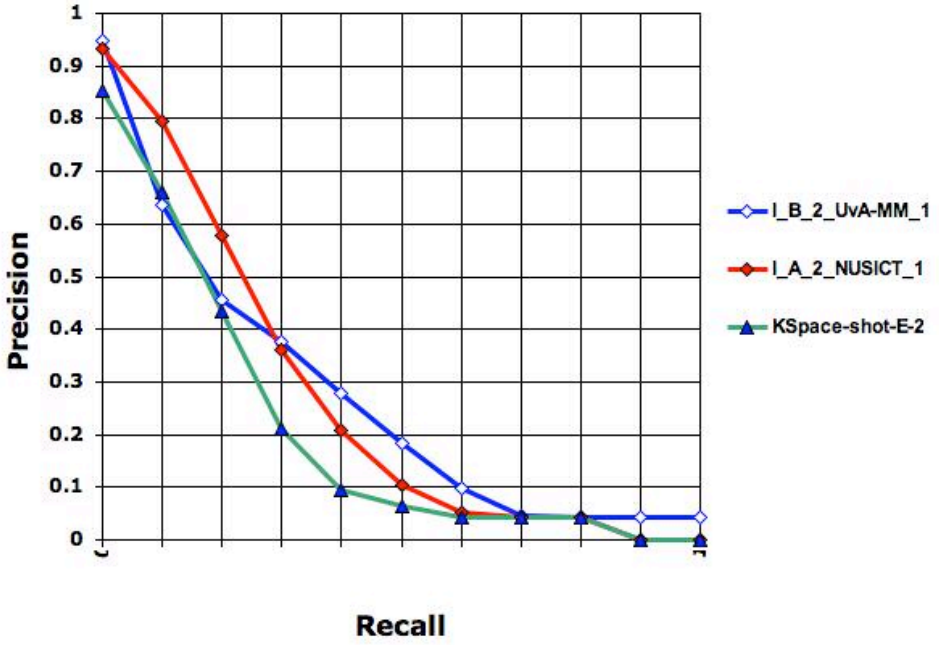
Figure 2: User interface for Dublin City University / K-Space Search System
452x285mm (72 x 72 DPI)

Figure 3: User interface for University of Amsterdam's Search System
152x113mm (72 x 72 DPI)

Figure 4: User interface for National University of Singapore's VisionGo Search System
150x108mm (72 x 72 DPI)

Figure 5: A sequence of multiple keyframes for shot213_62
163x26mm (72 x 72 DPI)

Figure 6: A sequence of multiple keyframes for shot149_62
163x26mm (72 x 72 DPI)



219x147mm (72 x 72 DPI)