

Manual and Interactive Search Runs at The University of Alabama and The University of North Carolina at Chapel Hill

Dan Albertson
University of Alabama
504 Gorgas Library, Box 870252
Tuscaloosa, AL 35487-0252
(205) 348-1517
dalbertson@slis.ua.edu

Javed Mostafa
University of North Carolina at Chapel Hill
CB #3360, 100 Manning Hall
Chapel Hill, NC 27599-3360
(919) 962-8366
jm@unc.edu

ABSTRACT

Researchers from the University of Alabama and the University of North Carolina at Chapel Hill collaborated on this study for TRECVID-08. This study focused on the search task of TRECVID-08, and the experiments included two full search runs, one interactive and one manual. Both search runs, M_C_2_ViewFinderALNC_2 and I_C_1_ViewFinderALNC_1, had similar designs. Each was conducted using the ViewFinder video retrieval system and employed the same expert user. The interactive submission, I_C_1_ViewFinderALNC_1, served as the baseline run for this year's search task (required for groups submitting runs involving human participants). As required by TRECVID-08 guidelines, the version of ViewFinder implemented for the baseline run (I_C_1_ViewFinderALNC_1) returned video shots using only the ASR¹/MT² output provided by the conference [5]. The ViewFinder system employed for the manual run, M_C_2_ViewFinderALNC_2, evaluated a QBE³ hybrid⁴ search. Differences across these two search runs, along with comparisons with the results from other participating groups, are reported. Results of this study indicated that significant progress is needed in order to become more competitive with some of the more notable research groups at TRECVID.

Keywords

Video retrieval, interface design, TRECVID, ViewFinder

1. INTRODUCTION

ViewFinder is an ongoing research project that explores user interaction and interfaces for video information retrieval (IR) systems. The ViewFinder system has been implemented for multiple experimental studies, including four years of TRECVID participation (2002-2004, 2008) and also a research project investigating interfaces for a Science Education video digital library. ViewFinder is intended to serve as a prototype system and to facilitate straightforward evaluation of certain interface features and functions across different contexts and datasets.

¹ ASR = Automatic speech recognition

² MT = Machine translation

³ QBE = Query by example

⁴ The hybrid QBE function implemented for this study searched using both color information and transcript terms.

For this year's workshop, ViewFinder was implemented to search and browse the Sound and Vision dataset using a variety of different techniques. There was a total of two search runs submitted, one interactive (I_C_1_ViewFinderALNC_1) and one manual, (M_C_2_ViewFinderALNC_2). Discussion of systems development, experimental methods, and results from these two search runs are presented in this paper.

After several years away from TRECVID, the researchers on this study used the 2008 Workshop primarily to get reacquainted with the forum. Even though this team previously participated, while at Indiana University, in years 2002-2004, this time back proved to be most challenging. The researchers of this study hope to use results from this year's Workshop to serve as a baseline for future years of TRECVID participation. Similar to previous years in TRECVID, this study applied a preexisting system, ViewFinder, to search and browse the test dataset and perform the search experiments.

2. BRIEF RESEARCH REVIEW

Considering the nature of the TRECVID workshop, it's necessary to review prior studies, including systems that were shown to be effective and demonstrated some uniqueness. More specifically, this section surveys features and functions for retrieving video that have been implemented and evaluated by both fellow TRECVID participants and others outside the Workshop. This discussion serves as a basic overview of some general features identified throughout various studies, and may be considered more suitable for readers who are new to video IR and/or TRECVID.

2.1 Interface Features and Functions

Systems developed for the TRECVID search task will obviously emphasize search and browse features. Systems from FX Palo Alto Laboratory (FXPAL), Dublin City University (DCU), IBM, and the University of North Carolina at Chapel Hill (UNC) are a few that are surveyed throughout this section [1][2][3][4][10][16].

Generally speaking, video search functions come in two different varieties: text- and content-based (i.e. visual). Textual searching can be implemented using a typical keyword search and shown to be effective when applied to video transcripts. Video transcripts are automatically extractable, come in the forms of ASR and closed captioning (CC), and typically comprise all words as they are spoken verbatim throughout a video. The systems surveyed here, DCU, IBM, FXPAL, and UNC, have all previously implemented and evaluated a transcript-based search feature for

their TRECVID systems, which has been a requirement for teams submitting search runs that involved human participants. Organizing video transcripts – or making them more useful for retrieval – has been performed using many different textual IR approaches. Variations of Okapi I and II and tf-idf weightings have been applied to video transcripts, along with the use of LSA⁵, and shown to be reasonably effective [2].

Textual searching of video systems is not limited to transcripts. The Open Video Project (UNC) previously incorporated a “detailed search” where users can query video descriptions, abstracts, titles, etc. from manually created Dublin Core records [9]. In addition, since manual indexing can be time consuming and expensive, existing software has been implemented to streamline the production of video metadata. IBM created one such video annotation software named VideoAnnEx, which imports digital video and provides a customized graphical user interface (GUI) for annotating and segmenting video clips [6]. VideoAnnEx includes many preset selections for annotating video, such as the presence of “man-made setting,” “fields,” “water,” “deer,” or even “robots.” This software also allows free text data entry where users can annotate video using the terms of their liking [6]. Annotations made in VideoAnnEx are exportable in XML.

Content-based (visual) searching has also been a major focus in video IR research. The most common content-based retrieval techniques include searching by color, texture, edge, shape, and different combinations of each, many of these features implemented as QBE searches. FXPAL’s TRECVID-07 system allowed users to append search queries with keyframes (selected from the video dataset) and prompt the system to return results based on color correlograms [2]. The Físchlár system (DCU) previously incorporated variations of color histogram and edge detection searches [3]. IBM’s TRECVID-07 system also made use of color histograms and correlograms, along with texture and edge measures [4].

Video retrieval systems also support browsing. Browse features can be designed to explore a video collection, i.e. “exploratory browsing,” browse within a particular video file, or scan search results. Open Video allows users to browse by keyword, date, genre, actor, and contributing organization [9]. On the other hand, McDonald and Tait (2003) evaluated color and serial browse features [10].

This discussion presented several general techniques for searching and browsing video. As the reader can imagine, the range of possible options for which all of these search and/or browse techniques can be incorporated into individual video retrieval systems is vast. To further illustrate this, one unique example from TRECVID-07 included FXPAL’s system, which evaluated collaborative video retrieval. FXPAL’s TRECVID-07 system allowed multiple users to perform different roles simultaneously throughout a video search experience [2]. For example, one user would be searching while the other user was browsing, and both would have the capability to view search histories, selected search results, and automatically suggested queries [2]. Also, IBM began investigating social tagging and annotation browsing in TRECVID-07 [4].

3. EXPERIMENTS

ViewFinder was implemented for the search task of TRECVID-08. Two full search runs were carried out, one manual and one interactive. ViewFinder was classified as a type ‘C’ system (system development is discussed in the next section). A total of two different ViewFinder variants were employed for this study (see Table 1 for description of both systems).

The interactive search run (I_C_1_ViewFinderALNC_1) fulfilled the baseline run required by conference organizers, intended for all studies involving human users (i.e. those performing either manual or interactive runs). This mandatory run evaluated shots returned using only the ASR/MT output provided by the conference; in other words, ViewFinder variant 1 was used for this run (see Table 1) [5]. A total of one subject (user) completed all 24 search tasks in sequential order; this subject was an “expert” user of the system and actually the project lead and developer for ViewFinder’s TRECVID-08 efforts. (The researchers of this study fully acknowledge that this type of interactive study isn’t in total concurrence with the “spirit” of the TRECVID interactive search task; however, these experimental methods had to be carried out in this manner due to time constraints.) Up to ten minutes was allowed for each topic and no other restrictions were placed on the subject in regards to predefined searching, query formulation, etc.

Table 1: System variants.

System	Evaluated Features and Functions
Variant 1	ASR/MT Search (Baseline Run)
Variant 2	Hybrid QBE (Supplemental Run)

A manual run (M_C_2_ViewFinderALNC_2) was performed next. The same human subject who carried out the interactive run, as just described, completed all 24 topics in order. However, for this search run, ViewFinder variant 2 (see Table 1) was employed. As a result, an additional feature could be tested, including the hybrid QBE (or Promote) search. Manual searches began with a relevant result from the video dataset and always executed a Hybrid Promote (QBE) search, which equally employed both color qualities and automatically extracted ASR/MT terms (from the “promoted” result) for the query. A maximum of ten minutes was allowed for each topic, although, substantially less time was needed.

Results were collected after each search topic was complete. A list of returned search results was formatted, according to the TRECVID submission requirements, and outputted by the system to a text file, once the user clicked the “Finish” button. Start and end times for each topic were manually recorded. The analyses, reported in this paper, comprise a variety of figures including precision at *n* shots, interpolated recall precision, mean of inferred average precision, number of relevant shots returned, and topic durations. Comparisons of many of these measures across the different search runs of this study, i.e. the manual and interactive search runs, are reported later in this paper, along with comparisons of runs from other groups.

⁵ LSA = Latent semantic analysis

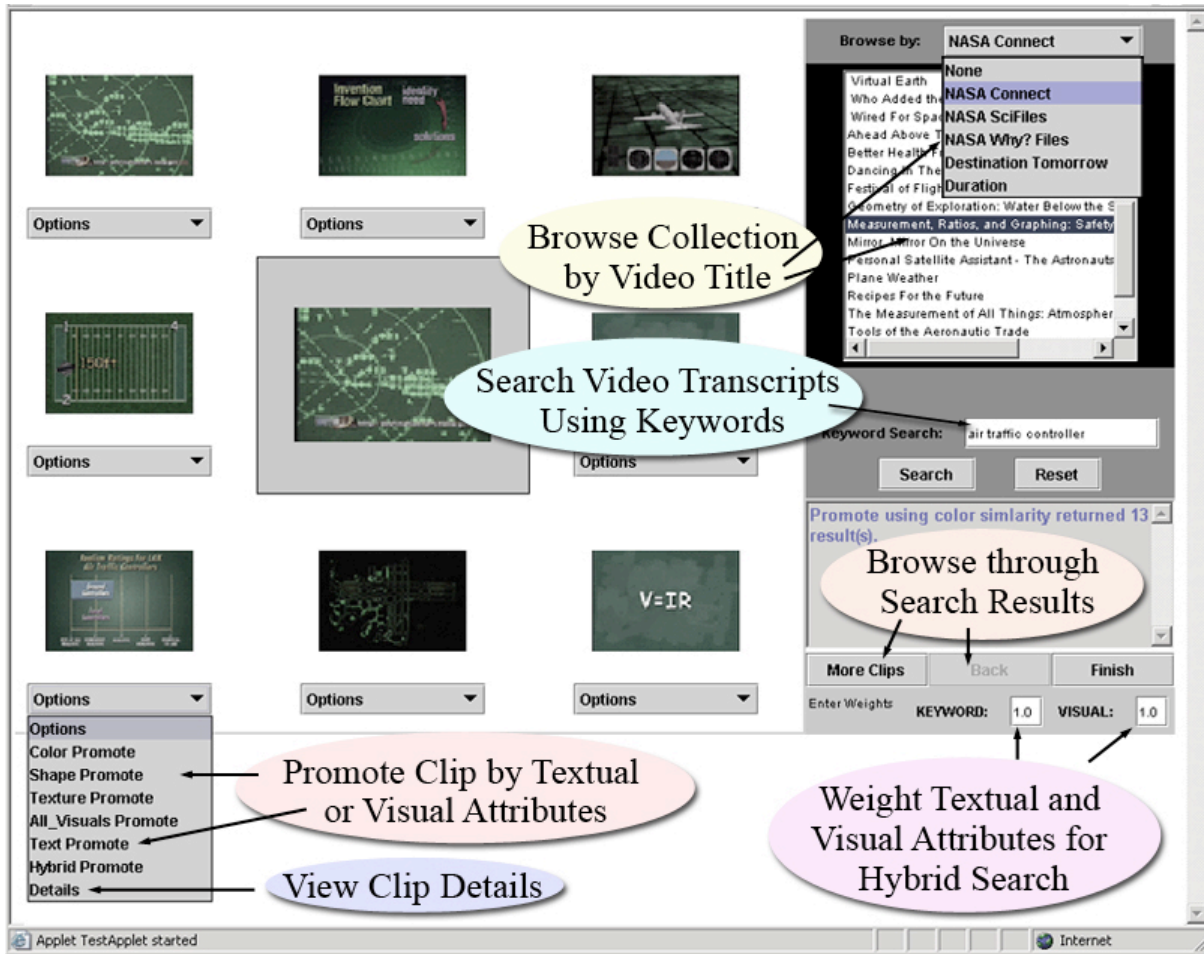


Figure 1: Screenshot of ViewFinder Video Retrieval System.

4. VIEWFINDER SYSTEM

ViewFinder is a web-accessible Java-driven search system supported by a backend Oracle database. Some of the Java-specific components of ViewFinder include the interface (see Figure 1) – developed using Swing – and Servlets running JDBC.

4.1 Dataset

Considering this round of TRECVID activity only involved participation in the search task, the dataset that was indexed and made searchable was the Sound and Vision video collection. This dataset was donated to the TRECVID community by The Netherlands Institute for Sound and Vision, and included approximately 100 hours of test video data comprising news magazines, science news, news reports, documentaries, educational programming, and archival video, all in MPEG-1 format. Only the test dataset and corresponding metadata (further described in the next section) were used for developing the TRECVID-08 experimental (ViewFinder) system.

ASR and MT outputs were provided to TRECVID research teams [5]. Because the Sound and Vision video was primarily Dutch programming, the ASR outputs were, subsequently, also in Dutch. Neither of the researchers on this study were fluent in speaking or reading Dutch, so the MT data became an important tool for developing this year’s search system. The MT data was a

Dutch to English translation of all ASR data extracted from the Sound and Vision dataset, transcribed and donated by Christof Monz of Queen Mary, University London. This MT data was indexed, and allowed for keyword searching and transcript-based QBE (Promote) functions. Keywords, or terms, deriving from the MT output, were also accessible to users through the “Details” feature.

Next, ViewFinder had to be implemented to return search results according to the common shot boundary references. Workshop organizers provided common shot boundaries to research teams [11]. Retrieving video according to these shot boundaries is a regular task for all years of the TRECVID search task and is necessary to ensure common evaluation. Shot boundary files also included unique shot ids and time stamps (start and stop times of individual shots) within video files [11].

Georges Quénot of CLIPS-IMAG generously provided the keyframes for the Sound and Vision test collection. The keyframes were similar to what was provided in recent years of TRECVID, including multiple keyframes for each shot with titles

containing “RKF⁶” and “NRKF⁷.” Having multiple keyframes per shot gave researchers from this study the option of choosing which individual image to designate for each shot. However, for this study, additional analysis was not performed on these keyframes in regards to selecting the best (or most representative) individual keyframe. The researchers of this study simply used the keyframe indicated by the “RKF” to represent each corresponding shot. For TRECVID-08, the researchers were only interested in the search tasks involving human participants, so common video processing tasks, such as keyframe extraction, wasn’t a top priority and – due to time constraints – couldn’t be performed in a timely enough manner. The researchers of this study are aware of common techniques for extracting keyframes, including using the middle frame (number) of a shot. Time stamps and shot ids associated with the keyframes were also provided along with the images so researchers could easily align keyframes with shots. This information was, again, supplied by CLIPS-IMAG.

Other metadata provided as part of this dataset corresponded to individual video files. This information (or metadata) for each video included a unique id, duration, source, filename, use, and format. This information is typically used to implement various browse capabilities of the ViewFinder system and can be tested in follow up interactive runs.

4.2 Indexing and System Architecture

Keywords derived from transcript data (ASR/MT) were used for retrieval. More specifically, each term from the MT output was indexed in an Oracle database and associated with a particular video and shot. The researchers of this study had to associate MT terms with shot ids by calculating the absolute difference between (MT) segment times and the start / stop times of each shot. After calculating this difference, the id for the shot with the least (or lowest) absolute time difference, in comparison with the MT segment times, was assigned to every term appearing in the segment being analyzed. The researchers are interested in learning more precise ways of associating keywords from MT files with shot ids, which appear across separate files and use different scales (and segmentations).

Transcript data was further organized by weighting each keyword. Tf-idf weights were modified⁸ to reflect a video’s structure and indexed alongside each transcript term in the Oracle database. Weighted transcript terms enabled retrieval on more precise levels of granularity and also allowed for ranking shots by relevance.

Other additions to ViewFinder – from past years of TRECVID participation – included incorporating color similarities for retrieval. Visual (or content-based) analysis of the keyframes extracted from the Sound and Vision data was performed using Oracle’s interMedia Java API [7]. A color similarity measure (i.e. score) was generated for each unique comparison of all keyframes in the collection. Significant scores were indexed in the Oracle

database along with the corresponding shot ids of the compared images.

4.3 Interface Features and Functions

Many interface features and functions have been created and tested across different experimental studies using the ViewFinder system (Table 2). ViewFinder has been implemented now for four years of TRECVID participation, a formal study investigating video retrieval for Science Education, and other projects. The appearance of these different interface features and functions across various selected studies is presented in Table 2. In addition, a screengrab of the graphical user interface (GUI) of ViewFinder, a Java Applet developed using Swing, is found in Figure 1.

Table 2: Interface features and functions from different versions of ViewFinder.

	TRECVID 2004 System	Sci Ed Experimental System (2007)	TRECVID 2008 System
Keyword Search of manual annotations, i.e. text		•	
Transcript Search	•	•	•
‘OR’ search by default	•	•	•
‘AND’ search function		•	•
‘NOT’ search function			
Phrasal searching			
Keyword Promote Search	•	•	•
Color Promote Search	•	•	•
Texture Promote Search		•	
Shape Promote Search		•	
All Visuals Promote Search		•	
Hybrid Promote Search	•	•	•
Adjustable weighting of Promote Search attributes	•	•	•
Browse by Date	•		
Browse by Source	•	•	•
Browse by Duration		•	•
Browse by Title		•	
Details feature	•	•	•
Use of Keyframes	•	•	•
Ranked search results (keyword weighting, visual qualities assessment, hybrid weighting)	•	•	•
Retrieval by shot	•		•
Retrieval by minute		•	
Retrieval by segment (manually segmented)			
Retrieval by video	•	•	•

⁶ RKF indicates the main keyframe identified for an individual shot by CLIPS-IMAG.

⁷ NRKF indicates additional keyframes extracted from a shot by CLIPS-IMAG.

⁸ IDF was computed using the number of shots per video and the number of shots where the terms appeared. TF represented the number of times the word appears in a shot.

ViewFinder’s interface consists of two panels. The search panel, on the right-hand side, contains the primary search and browse features of ViewFinder. A keyword search is available to query video transcripts, or MT outputs in this case. The browse options can be selected from the drop-down menu at the top (right) of the search panel. Different versions of ViewFinder over the years have allowed users to browse by title, source, date, and duration

(see Table 2 for specific browse features from different versions of ViewFinder).

The results panel, located on the left side of the interface, contains other search features, while also displaying keyframes for the search results; up to eight search results are displayed per page and ranked according to relevance⁹. The results panel also contains the Promote (QBE) and Details functions. Promote searches and the Details function can be selected from dropdown menus directly below each individual search result; each menu corresponds to the keyframe (search result) above it. All Promote searches execute a new QBE search. The TRECVID-08 version of ViewFinder incorporated a color, textual, and hybrid¹⁰ QBE search. The Hybrid Promote, which employed both color similarities and automatically extracted keywords from the MT outputs, was evaluated as part of the manual run and allowed the user to weight the textual and color attributes comprising the QBE query. The Details feature retrieved additional information about a particular search result, including title, transcript data, etc., and displayed it in a pop-up window.

5. RESULTS

Precision at n shots, interpolated recall precision, mean inferred average precision, and the number of returned relevant shots at various shot depths were measures generated by TRECVID assessors and distributed to all research groups [13]. Definitions for each of these measures can be further explored in the proceedings of TREC-10 [15]. Topic durations – or the amount of time the user took to complete each search topic – were also recorded and reported along with the search results. This section summarizes the collective results of each of these measures and then draws several comparisons across our manual and interactive search runs, and other runs submitted by TRECVID groups.

Table 3: Summary of manual search results.

Interpolated Recall Precision		Precision at n Shots	
0.0	0.7306	5	0.1500
0.1	0.0000	10	0.1000
0.2	0.0000	15	0.0666
0.3	0.0000	20	0.0584
0.4	0.0000	30	0.0472
0.5	0.0000	100	0.0316
0.6	0.0000	200	0.0192
0.8	0.0000	500	0.0172
1.0	0.0000	1000	0.0136

Composite results from the manual search run, including interpolated recall precision and precision at different shot depths, are presented in Table 3. In addition, a summary of ViewFinder’s overall performance on the manual run alone includes:

⁹ Top left to bottom right is the descending order of relevance.

¹⁰ Combining keywords and visual information for video IR is commonly referred to as hybrid retrieval.

- The manual search run returned a total of 164 relevant shots out of a total of 4,122 in the dataset, across all 24 search topics.
- ViewFinder scored a mean inferred averaged precision of 0.003 for the manual run.

Table 4: Summary of interactive search results.

Interpolated Recall Precision		Precision at n Shots	
0.0	0.4740	5	0.1666
0.1	0.0030	10	0.1084
0.2	0.0000	15	0.0944
0.3	0.0000	20	0.0792
0.4	0.0000	30	0.0666
0.5	0.0000	100	0.0516
0.6	0.0000	200	0.0324
0.8	0.0000	500	0.0224
1.0	0.0000	1000	0.0142

Results of the interactive run were also analyzed collectively. Readers can examine the interpolated recall precision and precision at different shot depths for the interactive run in Table 4: Summary of interactive search results. A concise list of other results reported for the interactive run includes:

- 170 relevant shots were returned out of the possible 4,122.
- The mean inferred average precision for the interactive run was 0.004.

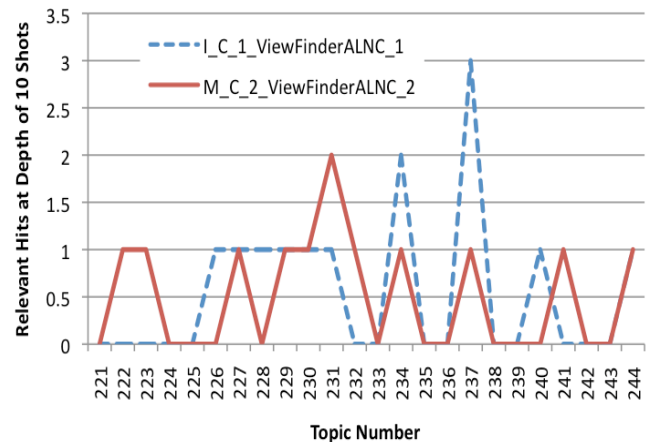


Figure 2: Relevant shots in top 10 of search results.

Results from these two search runs have been compared in order to detect any variations across the different run “flavors” and interface features being evaluated. As the reader can recall, the interactive run served as the baseline run – searching only the ASR/MT output – while the manual run evaluated results from a Hybrid Promote (QBE) search for every search topic. These runs have been compared using the analyses as just described,

including number of relevant shots returned, inferred average precision (for individual topics), and topic durations.

A comparison of the number of relevant search results returned across all individual topics is shown in Figures 2 through 5. Results are first presented (Figure 2) on how many relevant results were returned in the top 10 of the search results list. This set of results indicates that, over the course of all individual topics, the interactive run had a bit more success in getting relevant results in the top 10 search results than that of the manual run. The highest number of relevant results returned for an individual topic during the manual run at this depth was two (topic 231), while three was the highest score the interactive run (topic 237).

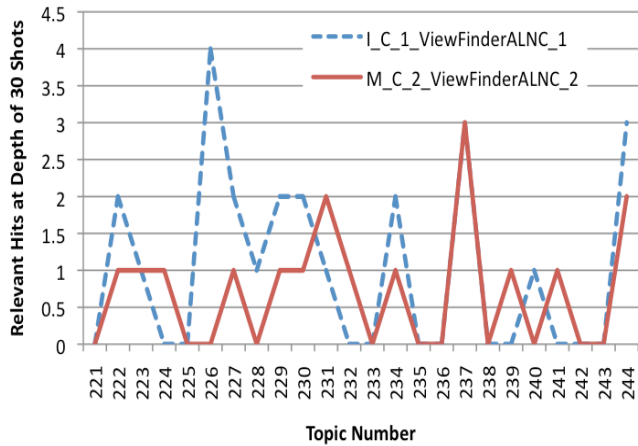


Figure 3: Relevant shots in top 30 of search results.

The same analysis was performed several more times, this time comparing the number of relevant results returned at depths of 30, 100, and 1,000 shots. These results are summarized in Figures 3 through 5. Here, readers can see how results from each run compared against the other in regards to returning relevant results at these depths.

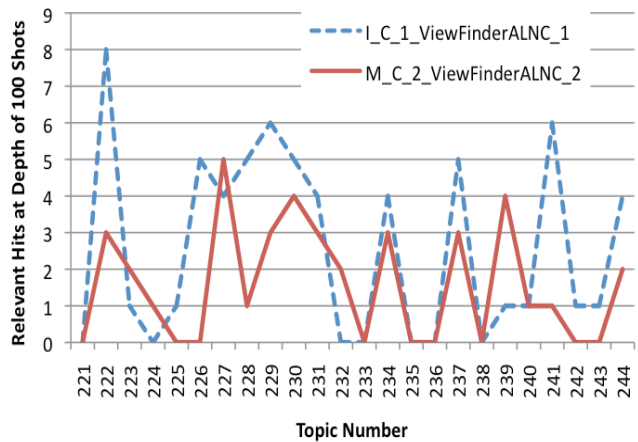


Figure 4: Relevant shots in top 100 of search results.

Inferred average precision across all individual topics of the manual and interactive runs are presented in Figure 6. Here, readers can see that the manual search run had an inferred average precision that ranged from 0.000 (on nine unique topics) to 0.011

(topic 223), while the range of inferred average precision for interactive topics was from 0.000 to 0.014 (topic 229).

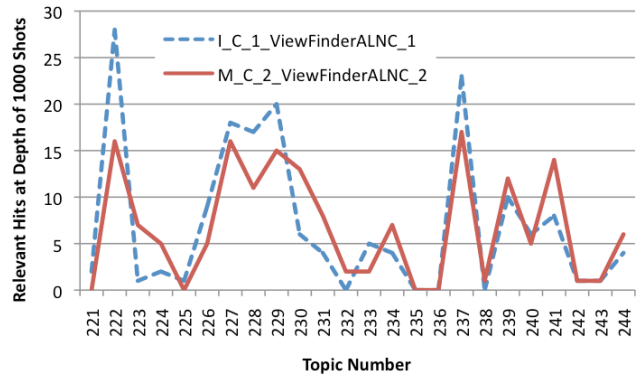


Figure 5: Relevant shots in top 1000 of search results.

The last measure used to compare differences between the manual and interactive runs is the recorded times (durations) for each topic. These results are presented in Figure 7. Obviously, since the design of interactive runs is intended for humans in the loop (search context) and the manual tasks consist of submitting one query, the times of the interactive topics are consistently greater than for the manual run. Times used to conduct each manual search ranged from 1.0 minute to 3.0 minutes, with an average of 1.60, and times for the interactive run ranged from 3.75 to 10.00 minutes (maximum allowed), averaging 8.08.

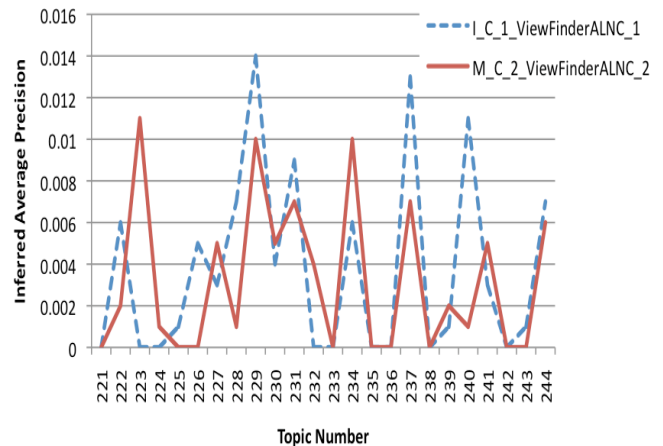


Figure 6: Inferred average precision of both runs.

Next, this study presents how the result for each of these two runs compared with other groups performing interactive and manual runs of the search task. These results are presented in Figures 8 and 9, and demonstrate that there were considerably more interactive runs than manual runs submitted for TRECVID-08. In Figure 8, the reader can observe how results from ViewFinder ranked according to all the other interactive runs. Results for the interactive run, ranked next to last among all other runs. However, remember that ViewFinder was employed only to perform one interactive run, which served as the baseline for this

year’s Workshop. Baseline runs only evaluated search results returned using the ASR/MT output.

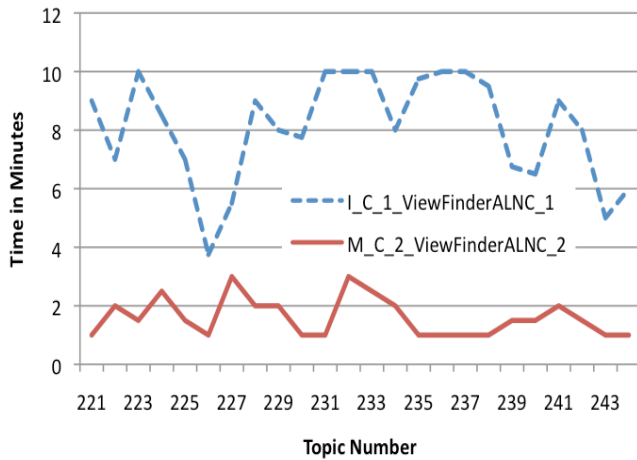


Figure 7: Time spent on topic completion for both runs.

The same comparison of mean inferred average precision for all submitted manual runs is also presented (Figure 9). This figure shows how results from ViewFinder ranked against all other manual runs. ViewFinder ranked last, but the queries analyzed in this run always included a hybrid QBE search, which incorporated a color search and “promoted” ASR/MT terms. The search experiments were designed this way to examine the effects of content-based and QBE retrieval for this year’s search task.

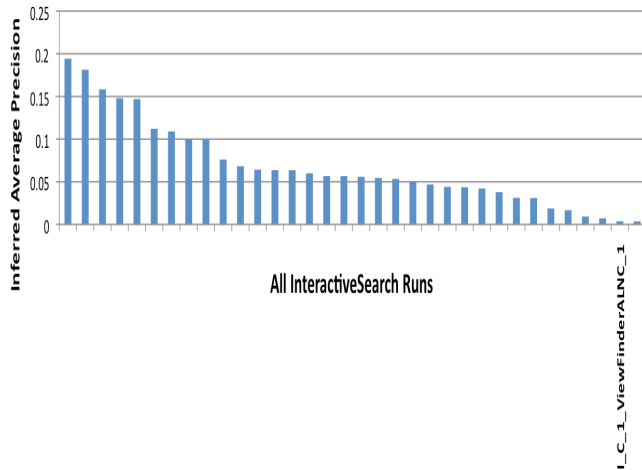


Figure 8: Mean of all topics in comparison with other groups.

While it would be possible to generate differences in inferred average precisions across individual topics and compare those scores across all the different runs from other groups, time restrictions prevented this analysis

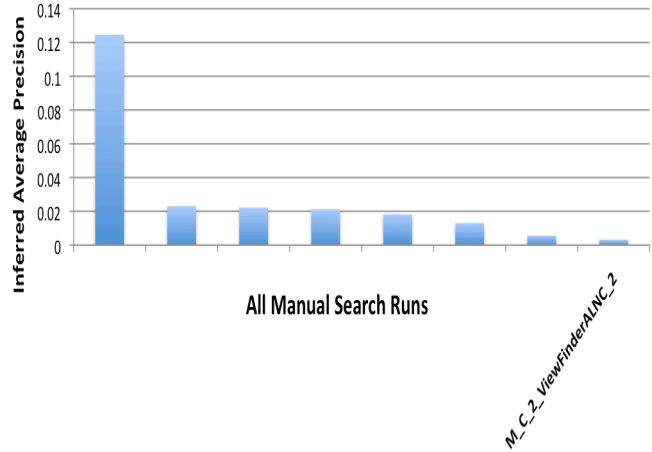


Figure 9: Mean of all topics in comparison with other groups.

6. DISCUSSION AND IMPROVEMENTS

The search experiments performed for TRECVID-08 included one manual and one interactive (baseline) run. Several observations can be made after reviewing the results from these experiments. These observations, as listed below, pertain to system improvements of ViewFinder and other general observations regarding influences affecting the outcomes of this study. Improvements to future experimental designs, including being more in agreement with the spirit of the TRECVID interactive search task, are being addressed by the researchers of this study. Several observations derived from this study include:

- The interactive search process, i.e. human in the loop, was shown to be an important factor. While results from the interactive run were regularly superior to results of the manual run, the interactive run incorporated fewer features (it only evaluated the keyword MT search).
- Results from the manual run did rival results from the interactive run. The manual run did outscore the interactive run – on a variety of measures – on a number of topics and only produced 6 less relevant shots across the entire run. This is significant because the user only got one attempt at submitting a query, which was always a hybrid QBE Promote search. Considering this, some merit for incorporating content-based (QBE) search functions into this search experiment may have been demonstrated.
- Following this study, the researchers intend to explore and test other ways for weighting transcript data. Incorporating an LSA approach for organizing transcript data is a goal for future years of TRECVID participation.
- The researchers of this study are interested in analyzing different techniques for clustering video shots in order to improve retrieval accuracy.
- ViewFinder needs to be more effective in limiting the overall number of returned search results. Since the researchers of this study are more interested in the search task, there’s a particular interest in improving the precision at higher depths of the search results.
- New approaches to content-based searching are always of interest to the researchers of this study. For future years,

the researchers of this study would like to incorporate additional content-based features, including shape, texture, and composite (visual) queries for retrieving video.

REFERENCES

- [1] Adams, B., Amir, A., Dorai, C., Ghosa, S., Iyengar, G., Jaimes, A. et al. (2003). IBM Research TREC-2002 Video Retrieval System. Proceedings of the Eleventh Text REtrieval Conference (TREC 2002), Gaithersburg, MD, November 19-22, 2002.
- [2] Adcock, J., Pickens, J., Cooper, M., Anthony, L., Chen, F., Qvarfordt, P. (2008). FXPAL Interactive Search Experiments for TRECVID 2007. Proceedings of the Text REtrieval Conference Video Retrieval Evaluation (TRECVID-2007), Gaithersburg, MD, November 5-6, 2007.
- [3] Browne, P., Czirik, C., Gaughan, G., Gurrin, C., Jones, G. J. F., Lee, H. et al. (2004). Dublin City University video track experiments for TREC 2003. Proceedings of the Text REtrieval Conference Video Retrieval Evaluation (TRECVID-2003), Gaithersburg, MD, November 17-18, 2003.
- [4] Campbell, M., Haurbold, A., Liu, M., Natsev, A., Smith, J.R., et al. (2008). IBM Research TRECVID-2007 Video Retrieval System. Proceedings of the Text REtrieval Conference Video Retrieval Evaluation (TRECVID-2007), Gaithersburg, MD, November 5-6, 2007.
- [5] Huijbregts, M., Ordelman, R., & Jong, F. (2007). Annotation of heterogeneous multimedia content using automatic speech recognition. Proceedings of SAMT, December 5-7 2007, Genova, Italy.
- [6] IBM Research - VideoAnnEx Annotation Tool. Retrieved April 15, 2004 from <http://www.research.ibm.com/VideoAnnEx/>
- [7] interMedia. Retrieved April 14, 2007 from <http://www.oracle.com/technology/products/intermedia/index.html>
- [8] Marchionini, G. (1995). Information seeking in electronic environments. Cambridge: Cambridge University Press.
- [9] Marchionini, G., & Geisler, G. (2002). The Open Video digital library. D-Lib Magazine, 8(12).
- [10] McDonald, S., & Tait, J. (2003). Search strategies in content-based image retrieval. Proceedings of the 26th Annual International Conference on Research and Development in Information Retrieval (ACM SIGIR '03) (pp. 80-87), Toronto, July 28-August 1, 2003.
- [11] Petersohn, C. (2004). Fraunhofer HHI at TRECVID 2004: Shot boundary detection system. Proceedings of the Text REtrieval Conference Video Retrieval Evaluation (TRECVID-2004), Gaithersburg, MD, November 15-16, 2004.
- [12] Shneiderman, B., Byrd, D., & Croft, B. (1997). Clarifying search: a user-interface framework for text searches. D-Lib Magazine, 3(1).
- [13] Smeaton, A. F., Over, P., and Kraaij, W. (2006). Evaluation campaigns and TRECVID. In Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval (MIR '06) (pp 321-330), Santa Barbara, California, USA, October 26 - 27, 2006, ACM Press, New York, NY.
- [14] Toms, E. G., Freund, L., Kopak, R., & Bartlett, J. C. (2003). The effect of task domain on search. Proceedings of the 2003 Conference of the Centre for Advanced Studies on Collaborative Research (CASCON '03) (pp. 303-312), Toronto, October 6-9, 2003.
- [15] Vorhees, E. M., & Harman, D. K. (Eds.). Common Evaluation Measures. (2001). NIST Special Publication 500-250: The Tenth Text Retrieval Conference, Gaithersburg, MD, A14 – A23.
- [16] Yang, M., Wildemuth, B., & Marchionini, G. (2004). The relative effectiveness of concept-based versus content-based video retrieval. Proceedings of the Twelfth Annual ACM International Conference on Multimedia (MM'04) (pp. 368-371), New York, NY, October 10-16, 2004.