

# The Use and Utility of High-Level Semantic Features in Video Retrieval

Michael G. Christel and Alexander G. Hauptmann

School of Computer Science, Carnegie Mellon University,  
Pittsburgh, PA, U.S.A. 15213  
{christel, hauptmann}@cs.cmu.edu

**Abstract.** This paper investigates the applicability of high-level semantic features for video retrieval using the benchmarked data from TRECVID 2003 and 2004, addressing the contributions of features like outdoor, face, and animal in retrieval, and if users can correctly decide on which features to apply for a given need. Pooled truth data gives evidence that some topics would benefit from features. A study with 12 subjects found that people often disagree on the relevance of a feature to a particular topic, including disagreement within the 8% of positive feature-topic associations strongly supported by truth data. When subjects concur, their judgments are correct, and for those 51 topic-feature pairings identified as significant we conduct an investigation into the best interactive search submissions showing that for 29 pairs, topic performance would have improved had users had access to ideal classifiers for those features. The benefits derive from generic features applied to generic topics (27 pairs), and in one case a specific feature applied to a specific topic. Re-ranking submitted shots based on features shows promise for automatic search runs, but not for interactive runs where a person already took care to rank shots well.

## 1 Introduction

Digital images and motion video have proliferated in the past few years, ranging from ever-growing personal photo collections to professional news and documentary archives. In searching through these archives, digital imagery indexing based on low-level image features like color and texture, or manually entered text annotations, often fail to meet the user's information needs, i.e., there is often a semantic gap produced by "the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation" [10]. Low-level features like histograms in the HSV, RGB, and YUV color space, Gabor texture or wavelets, and structure through edge direction histograms and edge maps can be accurately and automatically extracted from imagery, but studies have confirmed the difficulty of addressing information needs with such low-level features [6, 8]. This paper examines the use of high-level semantic features to assist in video retrieval and its promise to bridge the semantic gap by providing more accessible visual content descriptors. This paper explores that promise: are higher level

features like *outdoor*, *face*, and *animal* beneficial in news video retrieval, and can users correctly decide on which features to apply for a given need?

This paper focuses on the use of high-level features to overcome the semantic gap, and sidesteps the very real “sensory gap” problem for the video analysis community [10]. The sensory gap concerns the ease with which a person can infer a higher level feature like “car” in the scene, even if the car is profiled or partially occluded, and the relative difficulty to automatically detect “car.” A task within the TRECVID benchmarking community deals with evaluating the performance of automatic high-level feature detection [5]. Two criteria for selecting features to benchmark are “feasibility of detection and usability in real world semantic querying” [7]. We investigate the latter issue of usability by taking the first issue as solved: if we have features that are completely feasible to detect and can produce fully accurate feature classification, are these features useful for video retrieval?

Our concern regarding the promise of semantic features for bridging the semantic gap is motivated by user studies conducted with both the TRECVID 2003 (henceforth abbreviated TV03) and TRECVID 2004 (TV04) corpora. With a TV03 study, 13 users answered 24 topics with an interface supporting text search, image (color or texture) search, and browsing a “best” semantic feature set, e.g., best outdoor shots. In the study, text and image search accounted for 95% of the interactions with the semantic feature sets used only 5% [2]. Another TV03 study with 39 participants also found that high-level semantic features were hardly used in addressing the topics [4]. In a TV04 study, 31 users had access to the same query mechanisms: text search, color or texture image search, and browsing the semantic feature sets. The semantic feature sets were used only 4% of the time [1]. Why are the semantic feature sets not being used? One explanation is that the automatic feature classification is still too error-prone to be useful, a hypothesis we do not explore further here. It may also be that the users cannot decide which features apply to which topics, or that the inferred mapping of features to topics does not match the reality of the data, or that even with fully accurate feature classification the features do not address the topics well and would not improve topic retrieval. These latter questions are investigated by using the TRECVID features and search tasks from 2003 and 2004. We chose to work with TRECVID data because of the existence of pooled truth for both features and topics, leverage from prior TRECVID studies, the promise for follow-up repeatable experiments using published benchmarks by ourselves and others, and the noted enthusiasm by the TRECVID organizers for exploring the relevance of semantic features for retrieval, e.g., the 2004 overview report notes that “the ability to detect features is an interesting challenge by itself but it would take on added importance if it could serve as an extensible basis for query formation and search” [5].

TRECVID is an independent evaluation forum devoted to research in content-based retrieval of digital video [5]. The TRECVID test corpora for 2003 and 2004 was broadcast news from ABC, CNN, and (for 2003) C-SPAN, with 32,318 reference shots in the test video corpus for 2003 and 33,367 reference shots in 2004. The nontrivial size of the corpus, its definitions of sets of semantic features and information needs (topics), and human-determined truth for the features and topics provide a starting point for determining the utility of high-level semantic features for topic

retrieval, even though the chosen features were not always appropriate or comprehensive. The TRECVID topics include requests for specific items or people and general instances of locations and events, reflecting the Panofsky-Shatford mode/facet matrix of *specific*, *generic*, and *abstract* subjects of pictures [9]. A TRECVID overview categorizes all of the TV03 and TV04 topics into *specific* and *generic* (with no *abstract* topics) [5]. There were 8 *specific* and 16 *generic* TV03 interactive search topics; 7 *specific* and 17 *generic* TV04 topics considering only the 23 TV04 topics with confirmed answers in the corpus (topic 144 “Clinton with flag” was categorized as both).

The NIST assessors do not grade each of the reference shots for each of the topics and features, but instead grade the top  $x$  shots submitted by participants with  $x$  varying by topic and feature. Consistent with the NIST approach to relevance data, shots which were not scored manually were explicitly counted as not relevant to the feature or topic. By necessity, feature and topic descriptions are highly abbreviated here as we will investigate 638 feature-topic pairings, but they have unique IDs which can be used to look up complete descriptions on the TRECVID web site [5].

## 2 Data Analysis of TRECVID 2003 and 2004 Features and Topics

We began our analysis by examining the NIST human assessments of relevance of a shot to a feature and a topic. To get a sense which topics would benefit from which feature, we computed the probability of a shot being relevant to a topic  $P(S_t)$  and compared it to the probability of a shot being relevant to a topic, given the shot relevance to a particular feature  $P(S_t|S_f)$ . Rather than estimating chi-square correlation significance, we instituted a single threshold

$$P(S_t|S_f) - P(S_t) > 0.01$$

that filtered out both minimal absolute probabilities and minimal improvements in probability, which would be unlikely to substantially impact retrieval. We show in Figure 1 the features that can improve the chance of finding one or more topics by 1%. One side effect of our selection approach is that no negative features are found, explained by the imbalance of feature-relevant and feature-irrelevant shots. Only relatively few shots were judged relevant to a given feature, thus most shots were irrelevant [5]. In computing the negative feature, i.e.,  $P(S_t|S_{\sim f})$ , we are not reducing the overall search space much, resulting in likelihood improvements of less than 1%.

With 54 of the 638 possible topic-feature pairings showing benefit, based on estimates given the annotation “truth” about the features and topics, we were encouraged to investigate further. Also of interest was the pattern between *generic* and *specific* topics and features. Only 3 features were specific, feature 27 and 30 “Madeleine Albright” and feature 29 “Bill Clinton”, underlined in Figure 1. There are 7 *specific* feature to *specific* topic pairings, 39 *generic* to *generic* pairings, and only 8 between-class pairings (5 *generic* feature to *specific* topic, 3 *specific* feature to *generic* topic).

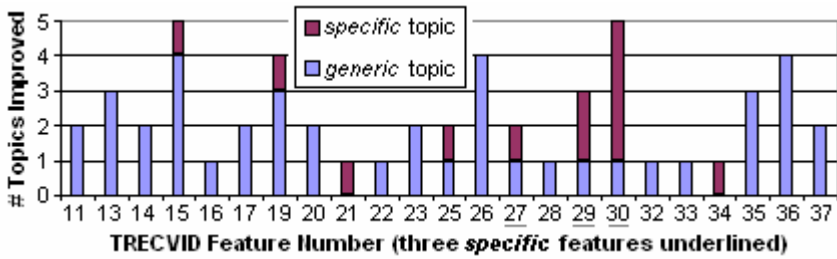


Fig. 1. Features improving one or more topics by > 1%, TV03 and TV04

### 3 User Study: Mapping Features to Topics

The purpose of the user study was to determine whether people associate semantic features to information needs uniformly. Twelve university employees and students participated in the study, with eight participants being very familiar with TRECVID features, topics, and video information retrieval and the remaining four relative novices. Each participant filled out two tables, one mapping 10 TV04 features to the 23 TV04 topics, and the other mapping the 17 features to the 24 topics in TV03. Each participant hence made  $230+408=638$  judgments as to the sign and degree with which a feature addresses a topic, which took from 40 to 80 minutes to complete. A total of 7656 human-generated judgments were produced in this way.

The survey set up the problem as follows: suppose there are tens of thousands of video shots and you need to answer a particular topic. You don't have time to look through all the shots, but can choose to either look at or ignore shots having a feature. For example, if the topic were "cherry trees" you might decide to definitely look at outdoor shots and vegetation, and definitely ignore Madeleine Albright shots. For each topic, rate whether each feature would help or hurt in finding answers to the topic according to this scale: *definitely ignore* the shots with the feature, *probably ignore*, *don't know*, *probably keep*, and *definitely keep* shots with the feature.

The judgments showed the difficulty in assessing the utility of a feature for a given topic. The overall correlation of ratings between pairs of subjects was weak for both TV03 and TV04. From the 66 pairings of raters, on TV03 the Pearson product moment correlation coefficient values ranged from 0.37 to 0.77, mean 0.58, STD 0.07. For TV04 the coefficient values ranged from 0.27 to 0.70, mean 0.56, STD 0.09. Hence, with both feature-topic sets there was weak positive correlation, but with coefficients too low to support the claim that a single human judge would represent human opinion on the relevance of a feature to a topic across all features and topics. Hence, approaches using human value judgments of feature relevance to topics are cautioned against reading too much into the value of a single judge.

The relevance of some features to topics is too ambiguous for people to express a clear, consistent opinion. For example, consider the feature "organized sporting event" and the topics regarding the Mercedes logo and snow peaks. For the Mercedes logo topic, ten subjects expressed an opinion aside from *don't know* with seven rating the feature as *definitely ignore*, one as *probably keep*, and two as *definitely keep*. For the snow peaks topic, for the eight subjects expressing an opinion, four rated the sporting event feature as *definitely ignore* while two rated it as *definitely keep* and two

others as *probably keep*. The broad nature of the “organized sporting event” feature made it more difficult to assess: subjects who thought the feature included auto races with prominent Mercedes logos would want to keep the sporting shots, and likewise subjects who thought skiing on mountain slopes might be included as sporting events would consider the feature relevant when looking for snow peaks. The collective evidence shows much disagreement between the raters across the feature-topic associations: over 20% of the TV03 associations and 17% of the TV04 associations have at least one rater scoring the association as *definitely ignore* while another rated it as *definitely keep*. By contrast, only 12% of TV03 associations and 13% of TV04 associations had strong uniformity of all 12 raters within one ratings point of each other on the 5-point scale. We are interested in looking at the feature-topic associations where raters did have greater agreement, i.e., those associations with ratings having a relatively low standard deviation. Tables 1 and 2 present the top quartile of associations having the best ratings agreement (which correlated to a standard deviation of < 0.7 for both sets). Empty cells in the tables indicate feature-topic associations having higher levels of disagreement amongst the 12 subjects.

**Table 1.** Average association on 5-point scale (1 = ignore, 5 = keep) from 12 raters assessing TV03 feature relevance to TV03 topics; blank cells indicate higher levels of rater disagreement, gray cells denote improvement according to truth data for topics and features from Section 2

	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
<b>Feature Key:</b>	14 building	18 female speech	22 non-studio setting														
	11 outdoors	15 road	19 car truck bus	23 sporting event	26 violence												
	12 face in news	16 vegetation	20 aircraft	24 weather news	27 Albright												
	13 3+ people	17 animal	21 speech in news	25 zoom in													
<b>100 aerial views</b>	4.9			4.8	4.8			2.9				4.6			3.0		1.2
<b>101 basketball</b>		2.8							1.4				5.0	1.3			1.1
<b>102 pitcher throw</b>	4.7						1.3						5.0				1.1
<b>103 Yasser Arafat</b>		4.9	3.9	3.2										1.1	3.2		
<b>104 airplane</b>	4.9								4.9						2.9		
<b>105 helicopter</b>	5.0				3.6				4.8		4.7						
<b>106 Tomb...Soldier</b>	5.0		3.6								4.8		1.4				
<b>107 rocket</b>	4.8				3.1						4.6		1.3	3.2			
<b>108 Mercedes logo</b>				4.5						2.6			1.4				1.4
<b>109 tanks</b>	4.8					3.5						4.8					
<b>110 diver</b>							3.3					4.7			3.2		1.0
<b>111 locomotive</b>	4.9						2.9						1.3				1.3
<b>112 flames</b>	4.7						3.2	3.1			4.7						
<b>113 snow peaks</b>	5.0					2.4	3.1				4.8			3.1			1.3
<b>114 bin Laden</b>		4.9					3.3				4.7		1.3	3.3			
<b>115 roads/cars</b>	5.0			4.9			3.3				4.8						1.2
<b>116 Sphinx</b>	4.9						3.0				4.8	1.2	1.4	2.9			1.4
<b>117 city crowd</b>	4.9		5.0	4.7			3.3				4.8			2.9			
<b>118 Mark Souder</b>		5.0					3.3										
<b>119 M. Freeman</b>				3.0			3.1								3.3		
<b>121 coffee cup</b>																	
<b>122 cats</b>						4.9	3.3								3.6		
<b>123 Pope</b>		4.9					1.4	2.8						1.4			
<b>124 White House</b>	4.9		3.5	4.9				3.3			3.5	4.7	1.2		3.2		

One other piece of information is shown in Tables 1 and 2: the cells corresponding to the 54 instances of topics improved by a feature as computed in Section 2 are shaded gray. Of 33 such TV03 topic-feature associations, 14 were found with high agreement by raters, 2 others were found by raters with high agreement but rated as *don't know* rather than *keep*, and 17 were rated with a variety of opinions. Of the 21 TV04 topic-feature associations, 6 were found with high agreement by raters but 15 were rated with a variety of opinions. When participants expressed a rating, the rating agreed with the pooled truth data. They missed expressing a clear consistent opinion on over half the shaded cells, though. Also, raters expressed additional consensus intuitions that were not supported by the pooled truth. Of the 43 topic-feature pairs in TV03 and 8 in TV04 marked with high agreement as positively associated ( $\geq 4$ ) in Tables 1 and 2, 29 of the 43 and 2 of the 8 were not substantiated by the pooled truth procedure of Section 2. Much of this omission can be traced back to known shortcomings of pooled truth and the assumption that ungraded shots are not relevant. Consider that 28 of the 29 unshaded cells with values  $\geq 4$  in TV03 concern features 11 (outdoors), 12 (face of person in the news), or 22 (non-studio setting). Users expected these features to matter, but the pooled truth did not confirm their relevance to topics, because the pooled truth sets for these features are likely too small. For example, 2429 shots were identified as non-studio setting shots, 7.5% of the news corpus, but non-studio shots probably constitute at least 15% of the corpus.

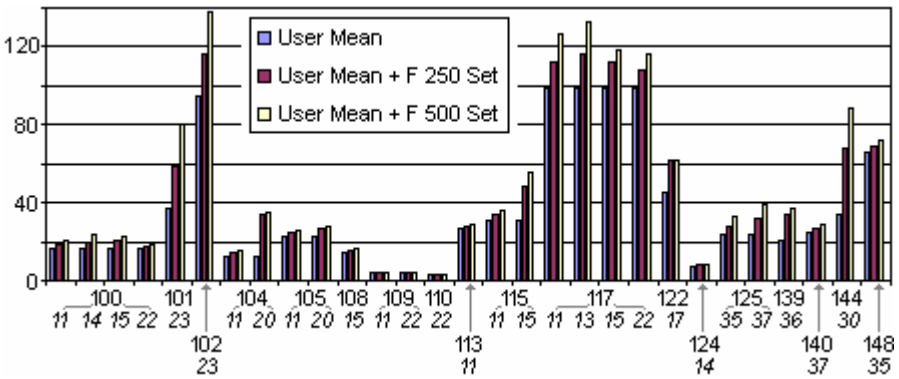
**Table 2.** Average association on 5-point scale (1 = ignore, 5 = keep) from 12 raters assessing TV04 feature relevance to TV04 topics; blank and gray cells same meaning as Table 1

TRECVID 2004	28 boat	29 Albright	30 Clinton	31 train	32 beach	33 basket	34 plane	35 people	36 violence	37 road
125 street	1.3					1.2		4.9		4.8
126 flood		1.1	1.3			1.0	1.3			
127 dog		1.1				1.0	1.1	4.8		
128 Hyde										
129 Dome	1.2				1.0	1.0				
130 hockey	1.0	1.2	1.2	1.2	1.0		1.0			1.3
131 keys	1.2	1.3	1.3			1.1	1.0			
132 stretcher		1.4	1.4				1.3			
133 Saddam						1.0				
134 Yeltsin						1.0		3.3		
135 Donaldson										
136 golf		1.3				1.2	1.0			
137 Netanyahu						1.0				
138 steps						1.1	1.3			
139 weapon						1.0			4.7	2.8
140 bike						1.0	1.3			4.6
141 umbrella						1.2				
142 tennis	1.4			1.3			1.2			
143 wheelchair										
144 Clinton			5.0			1.2				
145 horse						1.0	1.1			
147 fire		1.1	1.1			1.0				
148 sign						1.3	1.3	4.8		4.3

### 4 Using High-Level Features in Interactive Search

Consider the 43 topic-feature pairs in TV03 and 8 in TV04 marked as positively associated in Tables 1 and 2. If users had access to accurate feature classifiers, would they be useful for improving recall, to find relevant shots for the topic not located by text and image search strategies? We investigate this question using the pooled truth data from TV03 and TV04 and these 51 well-identified topic-feature associations.

To determine the magnitude of benefit from making use of these associations, we compared the success of a human interactive searcher on the topics with the potential offered by the features. Looking at the individual submissions to TRECVID, we noticed that the top variants of individual runs within a research group differed only minimally. We selected the best interactive search system by the top five groups for further analysis as representative of well-performing interactive video search systems. For each system on each topic, we are given the number of topic-relevant shots from the interactive searcher based on human assessment (pooled truth). For each feature strongly associated with the topic, we computed the number of additional shots relevant to both feature and topic, but not yet found by the user. Therein is the value of feature sets for interactive search: revealing additional relevant shots not found through other means. We averaged the top five systems’ performance within each topic to come up with the average count of relevant shots returned per topic by the user, shown as one bar labeled “User Mean” in Figure 2, and the average new shots introduced by a feature strongly associated with the topic.



**Fig. 2.** Count of topic-relevant shots for associated topic-feature pairs, showing mean user performance and boost provided with a set of 250 feature-relevant shots, and 500 feature-relevant shots

We need to account for varying feature set sizes, e.g., there are 258 TV03 shots with feature 20 “aircraft” but 2429 for feature 22 “nonstudio.” Of course showing the users all the nonstudio shots will show them topic-relevant shots because you are showing them 7.5% of the corpus! For our analysis, we assume that the topic-relevant shots are uniformly distributed within the feature-relevant shot set (i.e., if it

holds 5% topic-relevant shots, then a subset drawn from the feature set will still hold 5% topic-relevant shots). We make use of empirical evidence indicating that users can browse through 250 shots via storyboards and solve a visually oriented topic successfully in 4 minutes [3]. Another bar in Figure 2 shows the improvement in the topic-relevant shot count when the user has access to 250 feature-relevant shots. Given that the TRECVID topic time limit is set at 15 minutes, but accounting for user fatigue and leaving time for other inquiries, we also report statistics for when 500 (or all of them, if fewer than 500) feature-relevant shots are accessible.

Of the 51 user-identified strongly associated topics/features, 29 pass a threshold of introducing the user to at least 10% more topic-relevant shots (in a set of 500 feature-relevant shots) than the user had found. By this metric, topic-feature pair 127-35 (person walking a dog/people walking) does not pass the threshold, because the users averaged finding 25.2 of the 64 shots for topic 127, and letting them browse 500 of the 1695 feature-35 shots would only have brought in an average of 0.4 walking-dog shots not already in the user's set of 25.2. As another example, topic-feature 112-22 (flames/nonstudio) does not pass the threshold, as the users averaged 62.2 of 228 flame shots, and browsing 500 of the 2429 nonstudio shots would bring in an average of 4 more flame shots, less than a 10% improvement over the 62.2 already collected. The threshold eliminated 7 of the 14 associations with feature 11 "outdoors", all 4 of the associations with feature 12 "face", and 9 of the 13 associations with feature 22 "nonstudio." The other eliminated pairs were 127-35 and 148-37.

From the original set of 17 and 10 TV03 and TV04 features, 8 and 4 demonstrate value to interactive searchers if available to perfect levels of accuracy (remember we use pooled truth feature sets to focus the investigation on feature utility for topics, rather than assessing the current state of the practice for automatic feature classification). Topics receiving the most improvement have very strong and obvious associations with a feature: 101 and 102 (baseball, basketball) with 23 (sports), 115 with 15 (road traffic/roads), 122 with 17 (cats/animals), and 144 with 30 (Bill Clinton and flag/Bill Clinton). Topics 100 and 117 are interesting in all of their feature associations passed our threshold of 10% improvement (4 each), and that the broad nature of the topics (aerial building-road shots, urban crowds) saw improvement when a number of features are applied individually. Obviously, using the related features in combination can produce even greater benefits, underscoring the potential for high-level semantic features to address generic topics if users can easily intuit their applicability.

For 6 of 7 *specific* topics in TV04, feature associations either could not be identified or provided no additional value according to our threshold for interactive search. For 5 of the 17 *generic* TV04 topics, at least one feature provided a means to find additional topic-relevant shots. For 7 of 8 *specific* TV03 interactive search topics, feature associations either could not be identified or provided no additional value according to our threshold for interactive search. Of the 16 *generic* TV03 topics, 12 had at least one feature bringing in additional topic-relevant shots at our threshold levels. Thus, our interactive search run evaluation presents strong evidence that semantic feature sets, if capable of being produced to high levels of accuracy, will indeed benefit the user in addressing generic topics, but that specific topics are already addressed adequately through other means (such as text search against the narrative

text in news). For news video, users failed to identify or realize any benefit for features on 13 of 15 specific topics. The only exceptions were the feature Bill Clinton (30) when looking for Bill Clinton and the U.S. flag (144) – unusual in that the feature is specific rather than generic – and topic-feature 108-15 (Mercedes logo/road).

### 5 Impact of High-Level Features on Submitted Search Runs

In Section 4 we explored the use of features to introduce additional relevant shots for interactive search. Here, we look at an automatable strategy to re-rank ordered shot sets in search runs based on features. As in the interactive search analysis, for each submission category (interactive - I, manual - M, and, for TV04 only, automatic - A [5]) we selected the best system by the top five groups for analysis. For each of the ordered, submitted shots for a topic, if that shot had not been judged relevant for a particular feature, then the shot was moved to the bottom of the list. Thus the re-arrangement by feature grouped the submitted shots relevant to a feature at the top of the submission, and not feature-relevant at the bottom, and otherwise preserving the relative rankings. Alternatively, the absence of the feature was also tried. If the result of either re-ranking improved the average precision for a topic, then the ranking from the single feature which improved average topic precision the most was substituted for the original submitted ranking. The results using mean average precision (MAP) are shown in Figure 3 for the submission categories across TV03 and TV04.

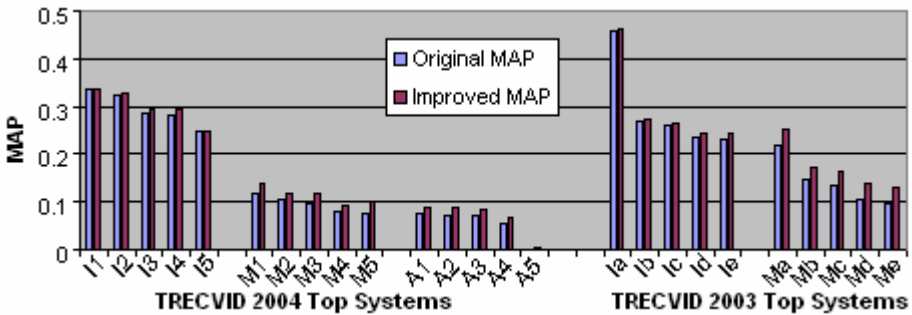


Fig. 3. Improvements in MAP from re-ranking submitted shots for a topic based on single feature, across submission categories for TV03 and TV04

Looking at TV04 results, the leftmost group is the interactive group, which essentially did not benefit from the re-ranking based on true features. The manual systems in the middle group all improved 12% to 30% relative to their original MAP. The final, fully automatic group, experienced improvements in MAP of at least 18%. Similar trends were found in TV03 submissions: negligible improvements for interactive runs, but, under optimal selection rules, noticeable improvements for non-interactive runs.

## 6 Conclusions and Acknowledgements

Pooled truth data gives evidence that some topics would benefit from features. A study with 12 subjects found that people often disagree on the relevance of a feature to a particular topic, including disagreement within the 54 of 638 positive feature-topic associations strongly supported by truth data (Figure 1). Mapping features to topics is not a trivial task, even with the small feature and topic counts in TV03 and TV04. If an interactive user makes the wrong judgment early in a retrieval session, e.g., choosing sporting events to solve the snow peaks topic but finding none because skiing is not covered, the user's frustration may drop to where feature browsing is no longer tried when addressing topics, one explanation for the lack of feature use discussed in Section 1. However, when multiple raters consistently agree that a feature is relevant to a topic, then the feature is likely to find additional topic-relevant shots.

The analysis of Section 4 found that for 29 of 51 topic-feature pairings, at least 10% more topic-relevant shots would be found by browsing 500 of the feature shots in well-performing interactive video search systems (19 more would provide benefits better than random, but not to the 10% level). Pooled truth for features 11, 12, and 22 is likely too small – if these features are excluded, then 18 of 20 topic/feature associations remain significant, an improvement over the stated ratio of 29 of 51. When topic-feature association is clear, topic performance can improve if users have access to ideal classifiers for those features. There is strong evidence that generic topics are helped more by features than specific topics, and that a specific feature only applies to a specific topic for which it is well associated. For a news corpus where some specific topic needs of the user community can be inferred (e.g., finding world leaders prominent in the corpus's time frame), developing "person X" feature classifiers tuned to those needs can produce a great performance improvement, as evidenced by our 144-30 pair. The improvement for specific features is limited to very few topics, however, and therefore a set of specific features will help little if the user has no correlated specific topics. Generic features, however, cut across a broader range of topics, and with ideal feature classification such features could present to the user many relevant shots for generic topics not retrieved by other means. Of the 29 topic-feature pairs in Figure 2, 27 are generic features applied to generic topics. Determining the features appropriate to a corpus and retrieval community's needs is important: only 12 of the 29 TV03 and TV04 features were clearly mapped to topics in our user study and then had confirmed benefits for interactive search (Figure 2). Finally, re-ranking submitted shots based on features shows promise for manual/automatic search runs, but not for interactive runs where a person already took care to rank shots well. For interactive runs, the potential improvement from features comes from introducing additional topic-relevant shots not found by other means. Future work will examine the impact of combinations of features to finding topic-relevant shots.

This material was made possible by the NIST assessors and the TRECVID community. It is based on work supported by the Advanced Research and Development Activity (ARDA) under contract number H98230-04-C-0406 and NBCHC040037.

## References

1. Christel, M.; Conescu, R.: Addressing the Challenge of Visual Information Access from Digital Image and Video Libraries. Proc. ACM/IEEE JCDL, ACM Press (June 2005)
2. Christel, M.; Moraveji, N.: Finding the Right Shots: Assessing Usability and Performance of a Digital Video Library Interface. Proc. ACM Multimedia, ACM Press (2004), 732–739
3. Christel, M., Moraveji, N., Huang, C.: Evaluating Content-Based Filters for Image and Video Retrieval. Proc. ACM SIGIR, ACM Press (July 2004), 590–591
4. Hollink, L., et al.: User Strategies in Video Retrieval: A Case Study. In: Enser, P., et al. (eds.): CIVR 2004. LNCS 3115. Springer-Verlag, Berlin Heidelberg (2004) 6–14
5. Kraaij, W., Smeaton, A.F., Over, P., Arlandis, J.: TRECVID 2004 – An Introduction. TRECVID '04 Proc., <http://www-nlpir.nist.gov/projects/tvpubs/tvpapers04/tv4overview.pdf>
6. Markkula, M., Sormunen, E.: End-user searching challenges indexing practices in the digital newspaper photo archive. Information Retrieval, 1 (2000) 259–285
7. Naphade, M.R., Smith, J.R.: On the Detection of Semantic Concepts at TRECVID. Proc. ACM Multimedia, ACM Press (2004), 660–667
8. Rodden, K., Basalaj, W., Sinclair, D., Wood, K.R.: Does organization by similarity assist image browsing? Proc. CHI '01, ACM Press (2001), 190–197
9. Shatford, S.: Analyzing the Subject of a Picture: A Theoretical Approach. Cataloguing & Classification Quarterly, 6 (Spring 1986) 39–62
10. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content based image retrieval at the end of the early years. IEEE Trans. PAMI, 22 (2000) 1349–1380