The Lowlands Team at TRECVID 2008

Robin Aly[1], Djoerd Hiemstra[1], Arjen de Vries[2], Henning Rode[2]
[1]University of Twente, The Netherlands [2] CWI, The Netherlands

February 24, 2009

Abstract

Type	Run	Description	MAP/mean infAP
	HLF Official		
H	utcwiprimw1-46	Our preliminary run for concept organization	0.0233
	Search Official		
F	utcwi-asr	ASR only run	0.0025
F	utcwi-abs	PRFUBE on 101 concept using Wiki abstracts	0.0037
F	utcwi-art	PRFUBE on 101 concept using Wiki articles	0.0034
F	utcwi-cuvro	PRFUBE on 374 columbia/vireo concepts using Wiki articles	0.0049
F	utcwi-vart	PRFUBE on 374 vireo concepts using Wiki articles	0.0093
I	utcwi-hand	PRFUBE with hand adjusted parameters	0.0040

In this paper we describe our experiments performed for TRECVID 2008. We participated in the High Level Feature extraction and the Search task. For the High Level Feature extraction task we mainly installed our detection environment. In the Search task we applied our new PRFUBE ranking model together with an estimation method which estimates a vital parameter of the model, the probability of a concept occurring in relevant shots. The PRFUBE model has similarities to the well known Probabilistic Text Information Retrieval methodology and follows the Probability Ranking Principle.

1 Introduction

The usage of a semantic representation of video objects through the occurrence of concepts is the prevalent search mechanism in today's Video Information Retrieval

(IR) search engines. Most current research aims at the creation of detectors for these concepts from low level features such as color histograms. Examples for these concepts are *Outdoor* or *Tennis*. The following search has to combine the output of the concepts in some way. We participated this year in the High Level Feature (HLF) or semantic concept extraction and the Search task.

For the extraction of concepts we followed the method from Diou et al. [4]. We trained a support vector machine (SVM) classifier from the manual positive and negative annotations. Our main interests were how to use the results of the noisy and sometimes faulty detectors.

Concepts either occur or are absent in video shots. In this way they are similar to the occurrence of words in Probabilistic Text IR, see [10]. In this paper we model the probability of the occurrence of a concept in relevant shots similar to the probability of a word occurring in relevant documents, which has been used for decades in Text IR. However, as the occurrence of a concept in a shot is not observable with certainty by a computer, we incorporate the probabilistic output of the predictions of the HLF extraction task.

This paper is structured as follows: In Section 2 we describe our work for the HLF extraction task and its evaluation. Section 3 describes the text only based search run and runs based on our novel search method. The conclusions in Section 4 and acknowledgments end this paper only followed by an appendix.

2 High Level Features Extraction Task

This year was the first time we participated in the High-level Feature Extraction task. We used a vector of 120 Weibull features as low level features which we extracted from the frame in the middle of the shot. The key frames had a resolution of $320x240$ pixels. The Extraction of the low level features is described in Diou et. al [4] which is using work from [11]. Furthermore, we use the Support Vector Machines (SVM) software package LIBSVM [3] with a C-Support Vector Classifier and a radial kernel function to detect the occurrence of concepts. For training we used the manual annotations from the collaborative annotation effort on this year's development corpus lead by Ayache and Quenot, see [2]. We propose following notation for the generated probability of the occurrence of a concept C given the extracted feature vector $\vec{F}$ and a SVM model θ_C: $P(C|\vec{F}, \theta_C)$. However, as we only used one model per concept we use here the shorter notation $P(C|\vec{F})$. The LIBSVM package estimates this probability according to Platt [8].

2.1 Model Optimization

We optimized the weights of the positive and negative class in the range of $[1..100]$ with steps of 10. The other parameters of the SVM were left to their default: $C = 1$ and $\gamma = \frac{1}{|\text{Shots}|}$. Instead of optimizing for classification accuracy we performed a three-fold cross-validation with the Mean Average Precision (MAP) as an optimization criterion. This way, models which rank shots with concept occurrences higher than others models were also preferred, even if none of the shots was classified to contain the concept, i.e.

Concept	Pos. Occ.	infAP	>Median
classroom	142	0.0090	
bridge	78	0.0019	
emergency_vehicle	32	0.0006	
dog	48	0.0029	
kitchen	152	0.0142	*
airplane_flying	56	0.0282	*
two_people	2698	0.0492	*
bus	45	0.0037	*
driver	197	0.0373	
cityscape	199	0.0320	
harbor	140	0.0035	
telephone	129	0.0070	
street	741	0.0469	
demonstration_or_protest	100	0.0047	
hand	1043	0.0423	
mountain	141	0.0233	
nighttime	283	0.0575	
boat_ship	326	0.0569	
flower	164	0.0364	
singing	222	0.0087	
MAP		0.0233	

Table 1: TV 2008 Concept Detections 'UTCWIPrimW1-46'

$P(C|\vec{F}) > 0.5$. For the cross-validation we randomly split the development collection in even parts with the same amount positive examples.

2.2 Performed Runs

We only created one run "utcwiprimw1-46". Unfortunately, due to a bug in our submission software, the feature identifiers were set incorrectly and our results could not be evaluated by NIST. Here, we present the corrected version of the run. Table 1 shows the inferred average precisions (infAP)of the detected concepts of our run. The mean infAP is 0.023. Out of 20 concepts we achieved in four concepts a better performance than the median among all evaluated systems. However, in general our results show that our extraction method still needs improvement.

To get an impression of the performance of our extraction method in other video domains and concept vocabularies we also trained two other set of models: 1) for the 36 official concepts from TRECVID 2007 and 2) for the 101 concepts of the MediaMill Challenge Set [9] based on annotations of the TRECVID 2005 development set. Table 2 shows the summary of the results on the mentioned data sets. The run on the TRECVID 2007 data showed similar performance to this year's official run. However, when evaluating the 101 concepts of MediaMill with the test subset of the

Concept Vocabulary	N^o Concepts	infAP
TRECVID 2007	36	0.0289
MediaMill	101	0.1990

Table 2: Summary of other runs

development set we get positive results. With a mean infAP of 0.1990 we are close the performance to the visual only extraction results from MediaMill (0.210 mean infAP). At the moment, the reason for this big performance difference is unclear to us. We plan to investigate this in the future. An overview of the single concepts is provided in the Appendix.

3 Search Task

In this section, we describe two distinct retrieval procedures. First, we describe the run only based on the Automatic Speech Recognition (ASR) output in Section 3.1. Second, we elaborate on our framework for binary unobservable events (PRFUBE) [1], which is described in Section 3.2. The following Section 3.3 describes the estimation for the probability of a concept occurring in a shot, which is an important parameter to the aforementioned retrieval model.

3.1 Automatic Speech Recognition based Search

For the text based run, we concatenated the one-best output of the speaker segments provided by Huijbregts et. al. [6]. If a shot lasted from t_1 until t_2 we included all speaker segments with $[t_{s1}, t_{s2}] \cap [t_1, t_2] \neq \emptyset$ where $[t_{s1}, t_{s2}]$ is the interval of the speaker segment. On average 2.6 speaker segments overlapped with a shot. Furthermore, we used the general purpose Text IR system PF/Tijah [5]. We only performed basic preprocessing on the text, removing all silence markers $[s]$ and used a standard snowball stemmer. The retrieval system PF/Tijah had the advantage that all queries could be executed from the provided topic XML file without any further modification in one execution.

3.2 PRFUBE

Our novel ranking framework PRFUBE is comparable to the Binary Independence Model in Probabilistic Text IR, see [10]. It estimates the probability of relevance, given a shot description of concept occurrences compared to a binary description of word occurrences in documents. However, due to the fact that the occurrences of the concepts are not observable by the computer, the ranking formula differs from standard text retrieval formulas. Note, we use an updated notation in comparison to Aly et al. [1].

$$P(R|S) \simeq P(R|\vec{F}) \propto \prod_{i=1}^{n} \Big[\underbrace{\frac{P(C_i|R)}{P(C_i)} P(C_i|\vec{F})}_{C_i \text{ occurs in shot}} + \underbrace{\frac{P(\bar{C}_i|R)}{P(\bar{C}_i)} P(\bar{C}_i|\vec{F})}_{C_i \text{ is absent in shot}} \Big] \quad (1)$$

Here, the ranking score is computed as follows: For each shot S, we observe a feature vector $\vec{F}$. The unobserved occurrences of n concepts are then used to calculate the probability of relevance given the detector output by marginalizing over their occurrences or absences. The probability $P(C_i|\vec{F})$ denotes the probability that concept C_i occurs in a shot given $\vec{F}$ (and θ_C). The value of this probability is generated by the SVM model using $\vec{F}$ as an input. The probability that a concept C_i occurs in relevant shots $P(C_i|R)$ is comparable to the probability of a word occurring in relevant text documents $P(w|R)$, which has been used in Probabilistic Text IR since decades. For the execution of the formula we need to estimate the probability $P(C_i|R)$ for each of the n concepts. An estimation method is provided in the following section. The probabilities at the right side of the summation (marked as "C_i is absent") can be derived from the values of the left side (marked as "C_i occurs") by subtracting the value from 1 (i.e. $P(\bar{C}_i|R) = 1 - P(C_i|R)$). The part of the formula marked with "C_i occurs" is equivalent to the Entropy based ranking formula proposed by Zheng et al. in [12]. However, their formula does not consider the case that a concept might be absent in relevant shots, which has been proofed vital for the performance of the search [1].

3.3 Query To Concept

In this section we describe a way to estimate the probability that a concept occurs in relevant shots $P(C|R)$ which is an essential parameter for the PRFUBE ranking model. The method uses existing training data which is annotated with concept occurrences to build a corpus of text representations of the annotated shots (here we used the LSCOM and MediaMill vocabulary from TRECVID 2005). The text representation for a shot is created by concatenating concept descriptions of the concepts which occur in this shot and the output of ASR. Ideally, a concept description meets two criteria: 1) it is precise (unambiguous) and 2) exhaustive, so that all words that a user could use to express his/her information need will be properly represented. We experimented with two different kinds of concept descriptions. Both descriptions contained the concept name and definition created as instructions for human annotators. Afterwards, we appended either 1) the Wikipedia Article discussing the concept or 2) the first 10 abstracts of Wikipedia articles returned by a search of the concept definition on the whole Wikipedia corpus. Wikipedia articles are known to contain a lot of noise while the abstracts are expected to be more precise but however might generate lower recall. The result of this procedure is a corpus of text documents which are subsequently indexed by any mature Text IR system.

At query time, the search engine first executes the textual query on the artificial text corpus. The result is a ranking of shots where each shot s of the development corpus has a score $score(s)$ attached. If there are r relevant documents in the development

corpus and knowing about the occurrences of the concepts annotation, the probability of a concept occurring given relevance is defined as:

$$P(C|R) = \frac{|C \cap R|}{|R|} = \frac{\sum_{i=1, s_i \in C}^{r} 1}{\sum_{i=1}^{r} 1} \tag{2}$$

Therefore, if the search engine gives reasonably good results we can assume a constant number of r relevant documents and calculate the estimate for the probability by the above formula. However, with a bigger r more and more irrelevant shots will be used in the estimation with equal influence as the shots from the top of the ranking which are more likely to be relevant. Therefore, we also investigate a method which takes the score of each shot into account:

$$P(C|R) = \frac{\sum_{i=1, s_i \in C}^{r} score(s_i)}{\sum_{i=1}^{r} score(s_i)} \tag{3}$$

The resulting estimation is in both variants properly normalized and can be used in the above described PRFUBE ranking model.

To see in how far a human is able to estimate the parameter $P(C|R)$ we let a user, who was slightly familiar with the data, make estimations for $P(C|R)$ for concepts for each official TRECVID 2008 query. Due to labor intensity we ranked the shots first by descending $P(C|R)$ calculated by using Equation 3 and asked the user only to judge the top 20 concepts. For each concept and query the user had to specify a value on a 6 point scale: one option for "ignore this concept" and one for following values of $P(C|R)$ $0\%, 25\%, 50\%, 75\%, 100\%$. An answer of x percent can be interpreted as follows: the "The concept occurs in $x\%$ of all relevant documents". We limited the choice to this scale because we believe that a user will not be able to judge a finer grained scale more objective.

3.4 Submitted Runs

Table 3 shows the results of our official runs. The first run is the ASR run which performed the worst. In all other runs we varied two dimensions as input into our PRFUBE ranking model. First, the concept and detector set which was used to produce the probabilities $P(C|\vec{F})$ and second the source of descriptions used for producing the artificial text corpus. As a concept detector set we used our models created by the methods described in Section 2 and the detector set from the joint work of Columbia University and Hong Kong University (VIREO) [7]. The text corpus was created from the MediaMill annotations on the development set of TRECVID 2005. For detectors from Columbia University and VIREO [7], which are based on the (constrained) LSCOM dictionary, we still used the MediaMill concepts for creating the textual shot representations and estimated the parameters of the LSCOM concepts based on the text run results of this corpus. This was done because experiments with the text corpus based on the LSCOM annotations produced worse results (possibly due to lower annotation coherence). We always used $n = 5$ concepts, which showed good results in the past.

As can be seen from the table all systems perform worse than MAP 0.01. We doubt these numbers are reasonably comparable. Mentionable is that the VIREO detectors

Name	Type	Concept Set	Desc. Kind	MAP	>Median
utcwi-asr	F	-	-	0.0025	0
utcwi-abs	F	MM101	Wiki Abstracts	0.0037	0
utcwi-art	F	MM101	Wiki Articles	0.0034	0
utcwi-cuvro	F	CU/VIREO	Wiki Articles	0.0049	4
utcwi-vart	F	VIREO	Wiki Articles	0.0093	7
utcwi-hand	I	TV07/08/MM101	-	0.0040	0

Table 3: Search Results TV 2008 Data / Type: F=Full automatic, I=Interactive / Concept Set: MM101=101 Concepts from MediaMill trained on TRECVID 2005, CU/VIREO, see [7], TV07=Official Concepts from TRECVID 2007, TV08=Official Concepts from TRECVID 2008 / Desc.: Type of description

alone with Wikipedia Articles for the estimation of the parameter $P(C|R)$ performed twice as good as all other runs. Both runs 'utcwi-cuvro' and 'utcwi-vart' answered four and seven respectively queries above the median. The run, where a human set the parameters $P(C|R)$ by hand, 'utcwi-hand', did not improve the performance either. However, our research questions about which concept set are helpful and what kind of concept descriptions were beneficial were not answerable.

Nevertheless, we assessed the described concept selection method, together with the according estimates. Table 6 in the Appendix shows the first five concepts selected for each query using Wikipedia Articles to build an estimation corpus. For space reasons only the queries which had to be answered by all search tasks are shown.

4 Conclusion

This year we participated in the HLF extraction task for the first time. The results of our detectors for both datasets from Sound and Vision (TRECVID 2007 and 2008) were around 2.00 mean infAP which we plan to improve in the future. However, for the TRECVID 2005 data our detectors showed a similar performance to the detector set from MediaMill detector set trained on visual features only which is a positive result.

Our search results were all beneath 0.01 MAP, which did not allow us to make further interpretations. We belief that the reason is the quality of the concept detectors which does not allow the search system to work properly. An informal assessment of the concept selection output shows that the estimations are plausible.

For next year, we plan to further intensify our efforts to build concept detectors . Furthermore, we will explore if we can more formally identify reason of the poor search performance.

5 Acknowledgments

We want to specially thank Christos Diou and his colleagues from CERTH-ITI for providing the software to extract the Weibull features from the key frame image. Also,

special thanks go to the Speech Group of our University who provided the ASR output. Furthermore, we would like to thank Stephane Ayache and Georges Quenot for coordinating the collaborative annotation effort. And last but not least, thanks to the teams from Columbia and Hong Kong University for providing their prediction output.

References

[1] Robin Aly, Djoerd Hiemstra, Arjen de Vries, and Franciska de Jong. A probabilistic ranking framework using unobservable binary events for video search. In *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 349–358, New York, NY, USA, 2008. ACM.

[2] Stéphane Ayache and Georges Quénot. Video corpus annotation using active learning. In *30h European Conference on Information Retrieval (ECIR'08)*, pages 187–198, March 30 2008.

[3] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[4] C. Diou, P. Panagiotopoulos, C. Papachristou, A. Delopoulos, M. Palomino, Y. Xu, and M. Oakes. Implementation of state of the art in cross-media indexing. Technical report, 2008.

[5] Djoerd Hiemstra, Henning Rode, Thomas van Os, Roel, and Jan Flokstra. Pftijah: text search in an xml database system. In *Proceedings of the 2nd International Workshop on Open Source Information Retrieval (OSIR), Seattle, WA, USA*, pages 12–17. Ecole Nationale Supérieure des Mines de Saint-Etienne, 2006.

[6] Marijn Huijbregts, Roeland Ordelman, and Franciska de Jong. Annotation of heterogeneous multimedia content using automatic speech recognition. In *Proceedings of the second international conference on Semantics And digital Media Technologies (SAMT)*, Lecture Notes in Computer Science, Berlin, December 2007. Springer Verlag.

[7] Yu-Gang Jiang, Akira Yanagawa, Shih-Fu Chang, and Chong-Wah Ngo. Cuvireo374: Fusing columbia374 and vireo374 for large scale semantic concept detection. ADVENT Technical Report #223-2008-1, Columbia University, 2008.

[8] J. Platt. *Advances in Large Margin Classifiers*, chapter Probabilistic outputs for support vector machines and comparison to regularized likelihood methods, pages 61–74. MIT Press, Cambridge, MA, 2000.

[9] Cees G. M. Snoek, Marcel Worring, Jan C. van Gemert, Jan-Mark Geusebroek, and Arnold W. M. Smeulders. The challenge problem for automated detection of 101 semantic concepts in multimedia. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 421–430, New York, NY, USA, 2006. ACM Press.

Concept	Pos. Occ.	infAP
airplane	20	0.0103
animal	531	0.0488
boat_ship	198	0.0359
car	481	0.0660
charts	63	0.0758
computer_tv-screen	293	0.0205
desert	24	0.0036
explosion_fire	47	0.0045
flag-us	10	0.0004
maps	100	0.0502
meeting	272	0.0484
military	299	0.0048
mountain	94	0.0309
office	1231	0.0299
people-marching	174	0.0154
police_security	165	0.0056
sports	323	0.0080
truck	118	0.0238
waterscape_waterfront	719	0.0809
weather	102	0.0142
MAP		0.0289

Table 4: TRECVID 2007 Concept Detections

[10] Karen Sparck-Jones, Steve Walker, and Stephen E. Robertson. A probabilistic model of information retrieval: development and comparative experiments - part 2. *Information Processing and Management*, 36(6):809–840, 2000.

[11] Jan C. van Gemert, Jan-Mark Geusebroek, Cor J. Veenman, Cees G. M. Snoek, and Arnold W. M. Smeulders. Robust scene categorization by learning image statistics in context. In *CVPRW '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, page 105, Washington, DC, USA, 2006. IEEE Computer Society.

[12] Wujie Zheng, Jianmin Li, Zhangzhang Si, Fuzong Lin, and Bo Zhang. Using high-level semantic features in video retrieval. In *Image and Video Retrieval*, volume Volume 4071/2006, pages 370–379. Springer Berlin / Heidelberg, 2006.

6 Appendix

Concept	Pos. Ex.	MAP LL	MAP MM	Concept	Pos Ex.	MAP LL	MAP MM
baseball	4	0.0139	0.0032	boat	249	0.0608	0.0956
hu_jintao	8	0.0204	0.0304	desert	250	0.0782	0.1029
sharon	13	0.0028	0.0497	natural_disaster	250	0.0318	0.0549
hassan_nasrallah	14	0.0016	0.0057	splitscreen	268	0.6370	0.6302
powell	14	0.0195	0.0102	cloud	270	0.1352	0.1174
clinton	15	0.0790	0.0037	grass	279	0.0963	0.0639
motorbike	16	0.0029	0.0061	flag_usa	285	0.1284	0.2273
tony_blair	20	0.0029	0.0051	police_security	286	0.0115	0.0116
waterfall	21	0.0082	0.3814	aircraft	306	0.0515	0.0725
beach	24	0.0218	0.0276	weather	307	0.4212	0.4049
swimmingpool	25	0.0010	0.0034	animal	309	0.1164	0.2094
candle	26	0.0093	0.0103	smoke	349	0.2710	0.2500
tank	26	0.0176	0.0084	maps	358	0.4716	0.4762
racing	27	0.0196	0.0289	truck	361	0.0351	0.0376
river	31	0.1114	0.3098	flag	390	0.0839	0.1892
dog	44	0.0669	0.2250	screen	475	0.0879	0.1005
nightfire	44	0.3523	0.5256	office	485	0.1057	0.0774
bird	56	0.6905	0.7236	mountain	508	0.1605	0.1405
cycling	57	0.0013	0.0422	soccer	517	0.5444	0.5030
football	61	0.0856	0.0477	waterbody	716	0.1383	0.1495
bicycle	63	0.0015	0.0061	corporate_leader	797	0.0170	0.0162
court	63	0.0841	0.0928	graphics	897	0.4261	0.3647
golf	78	0.2884	0.0908	monologue	962	0.1034	0.0942
duo_anchor	82	0.5110	0.6335	sports	1166	0.3000	0.3038
allawi	83	0.0003	0.0004	vegetation	1198	0.2223	0.1829
fish	83	0.4664	0.4890	military	1283	0.2182	0.2174
religious_leader	84	0.0360	0.0432	female	1359	0.0574	0.0856
government_building	85	0.0941	0.0106	meeting	1405	0.2429	0.2570
house	90	0.0224	0.0229	anchor	1578	0.6958	0.6309
kerry	91	0.0003	0.0004	male	1770	0.0917	0.0857
lahoud	93	0.2034	0.2886	building	2126	0.2551	0.3159
newspaper	97	0.5865	0.3753	vehicle	2360	0.2388	0.2212
prisoner	103	0.0065	0.0473	road	2404	0.2173	0.1947
tennis	105	0.3656	0.4483	violence	2500	0.3349	0.3168
fireweapon	108	0.0758	0.1215	government_leader	2899	0.1999	0.2130
snow	126	0.0574	0.0852	sky	3339	0.4741	0.4784
food	156	0.1933	0.2869	crowd	3559	0.4217	0.4802
explosion	164	0.0498	0.0981	urban	3651	0.2167	0.2217
chair	185	0.4049	0.4855	walking_running	4219	0.3615	0.3527
arrafat	193	0.0185	0.0257	studio	4234	0.6751	0.6358
basketball	217	0.2761	0.3817	indoor	6073	0.6084	0.5928
table	231	0.0594	0.0727	entertainment	6088	0.1854	0.1657
tower	231	0.0434	0.0570	outdoor	10130	0.7359	0.6879
charts	234	0.2440	0.3273	overlayed_text	11261	0.5903	0.6691
tree	241	0.1102	0.1243	face	19883	0.8909	0.8949

Table 5: TV 2005 Concept Detectors / LL=Low-Lands Team - us / MM=MediaMill Visual Only

QID	Query Text Concept $P(C\|R)$
0221	**Find shots of a person opening a door** people 0.92, face 0.90, studio 0.80, indoor 0.80, splitscreen 0.80
0222	**Find shots of 3 or fewer people sitting at a table** table 1.00, people 1.00, indoor 1.00, meeting 0.94, face 0.92
0223	**Find shots of one or more people with one or more horses** horse 0.98, people 0.88, animal 0.84, horse_racing 0.71, sports 0.67
0224	**Find shots of a road taken from a moving vehicle, looking to the side** vehicle 1.00, outdoor 0.92, road 0.92, overlayed_text 0.61, car 0.47
0225	**Find shots of a bridge** outdoor 0.98, tower 0.96, building 0.96, sky 0.75, urban 0.55
0226	**Find shots of one or more people with mostly trees and plants in the background; no road or building visible** tree 1.00, outdoor 0.90, building 0.76, sky 0.55, vegetation 0.39
0227	**Find shots of a person's face filling more than half of the frame area** outdoor 1.00, vehicle 1.00, bicycle 1.00, sports 0.88, cycling 0.88
0228	**Find shots of one or more pieces of paper, each with writing, typing, or printing it, filling more than half of the frame area** newspaper 0.51, studio 0.51, indoor 0.51, drawing_cartoon 0.49, cartoon 0.49
0229	**Find shots of one or more people where a body of water can be seen** outdoor 1.00, waterbody 1.00, sky 1.00, people 0.41, face 0.31
0230	**Find shots of one or more vehicles passing the camera** screen 0.84, entertainment 0.26, vehicle 0.16, bus 0.14, overlayed_text 0.12
0231	**Find shots of a map** graphics 1.00, maps 1.00
0232	**Find shots of one or more people, each walking into a building** people 1.00, people_marching 0.92, building 0.80, crowd 0.73, outdoor 0.71
0233	**Find shots of one or more black and white photographs, filling more than half of the frame area** studio 0.96, indoor 0.96, people 0.96, face 0.14, overlayed_text 0.14
0234	**Find shots of a vehicle moving away from the camera** entertainment 1.00, vehicle 0.02, boat 0.02
0235	**Find shots of a person on the street, talking to the camera** people 1.00, crowd 0.75, face 0.49, entertainment 0.35, police_security 0.31
0236	**Find shots of waves breaking onto rocks** people 0.65, outdoor 0.63, waterbody 0.63, beach 0.63, sky 0.51
0237	**Find shots of a woman talking to the camera in an interview located indoors - no other people visible** people 1.00, face 1.00, hassan_nasrallah 0.61, indoor 0.41, female 0.39
0238	**Find shots of a person pushing a child in a stroller or baby carriage** people 1.00, tony_blair 0.98, face 0.96, overlayed_text 0.41, government_leader 0.39
0239	**Find shots of one or more people standing, walking, or playing with one or more children** people 1.00, entertainment 1.00, face 0.86, overlayed_text 0.76, monologue 0.65
0240	**Find shots of one or more people with one or more books** people 1.00
0241	**Find shots of food and/or drinks on a table** table 1.00, people 1.00, indoor 1.00, meeting 0.94, face 0.90
0242	**Find shots of one or more people, each in the process of sitting down in a chair** chair 1.00, people 0.98, face 0.77, indoor 0.53, meeting 0.49
0243	**Find shots of one or more people, each looking into a microscope** people 0.71, animal 0.69, overlayed_text 0.65, walking_running 0.33, outdoor 0.27
0244	**Find shots of a vehicle approaching the camera** entertainment 1.00, vehicle 0.75, people 0.57, car 0.51, face 0.47

Table 6: Automatic Concept Selection for TRECVID 2008 topics 0221-0244