ISSN:2348-2079

Volume-5 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

Multimodal fusion for video search re ranking

[1]Mr R.Navinkumar MCA., M.Phil., Assistant Professor,
[2]Mr T.Sathishkumar Final MCA,
Department of MCA, Nandha Engineering College (Autonomous), Erode-52.
E-Mail ID: Navinsoccer07@gmail.com, sathishkumar6887@gmail.com

Abstract - Most of the users pay large attention only in top ranked portion of returned search results .So it is very essential to achieve high accuracy on top ranked documents there are so many methods to boost video search performance, they either pay less attention to the above factor or encounter difficulties in practical applications. In order to develop retrieval effectiveness, We should present This paper the on a flexible and effective re ranking method, called CR-Re ranking. For the purpose of fusing multimodal cues CR-Re ranking employs a cross-reference (CR) it is used to offer high accuracy on top ranked results Specifically, multimodal features are first utilized separately to re rank the initial returned results at the cluster level, and then all the ranked clusters from different modalities are cooperatively used to infer the shots with high relevance. Experimental results show that the search quality, especially on the top-ranked results, is improved significantly

Index Terms—Clustering image/video retrieval, multimedia databases.

1 INTRODUCTION

As an emerging research field, content-based video retrieval (CBVR) has attracted a great deal of attention in recent years. While various recovery models have been developed to prove video search quality, most of them implement search procedure by implicitly or explicitly measuring the similarity between the query and database shots in some low-level feature spaces. However, such similarity is not usually consistent with being perception due to the limitation of current image/video understanding techniques. That is, the semantic gap exists between the low-level features and high-level semantics. For example, although a scene with red flags and a scene with red buildings share similar color features, they have completely different semantic meanings. The semantic gap will enlarge linearly with the increase of data set size since a larger data set means more confusion, which thereby leads to rapid deterioration of search performance. Performance comparison between TRECVID'05 and TRECVID'06 evaluation on all the three search types, automatic, manual, and interactive, also reveals it. Consequently, it is more attainable for low-level features to reliably distinguish different shots in a relatively small collection, which is the basis of proposed re ranking scheme.

If we consider that the final aim of search engines is to meet users' information needs, it is reasonable to take user satisfaction and user behavior into account when designing a search engine. According to the analysis in, users are rarely patient to go through the entire result list. Instead, they usually check the top-ranked documents. Most of user click-through data from a very large Web search engine log also reflects such preference. Therefore, it is more essential to offer high accuracy on the top-ranked documents than to improve the whole search performance on the entire result list [5].

1.1 Related Work

Various methods have been planned for improving the retrieval performance of video search engines. The earlier work, which is based on relevance feedback (RF) strategy focuses mainly on the refinement of the initial search results in an interactive fashion. However, to use RF-based methods require users' labeling for updating the query model, which is usually time-consuming and

even impractical in some search scenarios. In contrast, pseudo relevance feedback (PRF)-based methods assume that the top-ranked documents are relevant and use them to automatically refine the search process [12]. For request, the co retrieval algorithm [13] treats the top-ranked results as positive examples and others as negative ones. Using these noisy training samples, a retrained retrieval model is then built via an Ad a boost-based ensemble learning method. Although both RF- and PRF-based methods have achieved precision improvement on the entire result list by returning more relevant shots, no mechanism guarantees that these relevant shots will be top positioned. The meta search strategy, which is initially put forward in the field of information retrieval, is imported to CBVR for improving video retrieval effectiveness. The input idea of meta search is that many result lists returned by several different search engines in response to a given query are aggregated into a single list in an optimal way. Meta search is generally based on the "imbalanced overlap property": different search models retrieve many of the same relevant documents, but different irrelevant documents. Using this property, the combination of the returned lists is performed by simply giving higher ranks to the documents that are contained simultaneously in multiple result lists. Similar schemes include the PageRank-like graph-based approach and the model-based re ranking algorithm.

In addition, it is not easy in practice to get access to multiple search an alternative scheme, the re ranking method can improve search quality by reordering the initial result list. Although the total number of relevant documents remains fixed after re ranking, the accuracy improvement at the low depth of the result list can be expected by forcing true relevant documents to move forward. It finds some relevance-consistent clusters first and then ranks shots within the resulting clusters. In this method, however, multiple modalities are integrated in a unique feature space, that is, multimodal features are fused by concatenating them into a single representation. This fusion strategy is called early fusion. As a consequence, IB-Re ranking is carried out only in a single feature space by which the accuracy on the top-ranked documents receives relatively less attention. Particular mention should be made to Kennedy et al.'s work, where a similar structure to is smartly exploited to build a vivid pictorial map of the world from the user-shared multimedia resources.

1.2 Basic Idea of CR-Re ranking

In this paper, we introduce are ranking method, called CR-Re ranking, which combines multimodal features in the manner of cross reference. The fundamental idea of CR-Re ranking lies in the fact that the semantic understanding of video content from different modalities can reach an agreement. Actually, this idea is derived from the multi-view learning strategy a semi supervised method in machine learning. Multi view learning first partitions available attributes into disjointed subsets (or views), and then cooperatively uses the information from various views to learn the target model. Its theoretical foundation depends on the assumption that different views are compatible and uncorrelated In our context, the assumption means that various modalities should be comparable in effectiveness and independent of each other. Multi view strategy has been successfully applied to various research fields, such as concept detection. However this strategy, here, is utilized for inferring the most relevant shots in the initial search results, which is different from its original role. CR-Re ranking method contains three main stages clustering the initial search results separately in diverse feature spaces, ranking the clusters by their relevance to the query, and hierarchically fusing all the ranked clusters using a cross-reference strategy.

1.3 Our Contributions

In our work, three key assistances are made to the video search re ranking. The first contribution is that multiple modalities are considered individually during clustering and cluster ranking processes. It means that re ranking at the cluster level is conducted separately in distinct feature spaces, which provides a possibility for offering higher accuracy on the top-ranked documents. In contrast, in previous work, multimodal features are first concatenated into a unique feature, and the consequent clustering and cluster ranking are then implemented once in the above unique feature space. The second contribution is defining a strategy for selecting some query-relevant shots to convey users' query intent. Instead of directly treating the top-ranked results as relevant examples like PRF, we further filter out some

irrelevant shots using some properties existing in the initial rankings. Reliably selecting a query-relevant shot set has a beneficial effect on cluster ranking. A cross-reference strategy the third contribution is presenting to hierarchically combine all the ranked clusters from various modalities. We assume that the shot with high relevance should be the one that simultaneously exists in multiple high-ranked clusters from different modalities.

Based on this assumption, the shots with high relevance can be inferred cooperatively using the cross-reference strategy and then be brought up to the top of the result list. As a result, the accuracy on the top-ranked documents is given more consideration. Because the "unequal overlap property" is employed implicitly, this fusion strategy is similar to the meta search methods to a certain extent. However, our cross-reference strategy differs in two ways from meta search. The first difference is that, instead of combining multiple ranked lists from different search engines, we integrate multiple reordered variants of the same result list obtained from only one text-based video search engine. The second one is that, instead of fusing multiple lists at the shot level, we first coarsely rank each list at the cluster level, and then integrate all the resulting clusters hierarchically Experimental results indicate that CR-Re ranking method indeed achieves higher accuracy on the top-ranked shots.

1.4 Organization

The rest of the paper is organized as follows: A comprehensive analysis on both the weakness in current video search engines and the feasibility for alleviating it. Then elaborates the proposed CR-Re ranking scheme. Experimental results and performance analysis are given in detail.

2. PROBLEM ANALYSIS

Currently, text information associated with video content is the main source used in successful semantic video search engine. In those search engines, researchers give much consideration to feature extraction and similarity measurement. Before presenting the proposed re ranking scheme, in this section, we first analyze the weakness in those search engines and then judge whether it is possible to alleviate the weakness using the re ranking technique.

2.1 Weakness of Current Search Engines

As a well-recognized community for video search, NIST TRECVID provides 24 query topics for all participants to test their video search systems. In annual competition, all

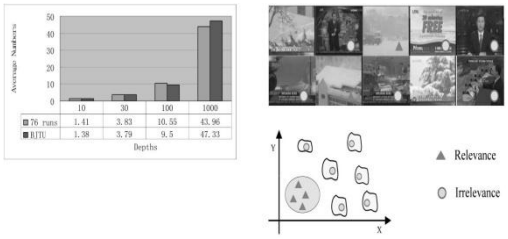

Fig.2. Diagrammatic sketch of centralization and decentralization properties in a 2D visual feature space.

In brief, current text-based video search engines generally cannot satisfy users well; it is necessary and possible to improve the search quality by performing an effective re ranking procedure.

2.2 Feasibility for Alleviating the Weakness

In addition, some observations on the initial rankings are helpful in building an effective re ranking scheme. Browsing over the top-30 results of all the 24 initial rankings in BJTU run shows that the true relevant shots are usually similar in view of visual perception, yet irrelevant ones are significantly different from each other. We call them centralization attribute of the relevant shots and decentralization property of the irrelevant shots, respectively. Although the relevant shots are scarce at low depths, there is a relatively large number of relevant shots at some great depths (e.g., depth ¼ 1; 000, average number ¼ 43:96). Therefore, it becomes feasible to boost the search precision at low depths by forcing those relevant shots at great depths to move forward. In other words, it is practicable to improve the accuracy of the top-ranked shots by reordering the initial search results. we can obtain three clusters from each feature space, which are needed for the hierarchical fusion in the following steps.

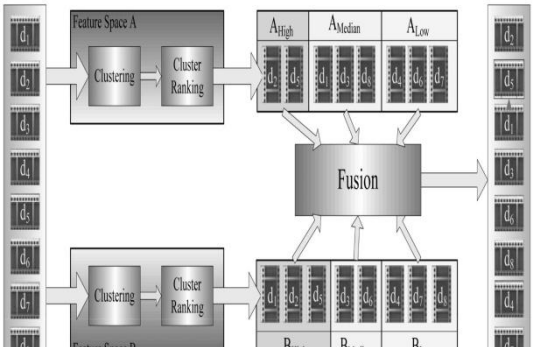

Fig. 3. Framework of proposed CR-Re ranking method.

As mentioned previously, low-level features are more suitable for discriminating different shots within a finite shot set. In our case, the initial result list of 1,000 shots used for re ranking is a relatively small shot set. Hence, it is possible to nicely partition the initial list into several clusters in certain low-level feature spaces. Specifically, after extracting multiple features for each shot, we carry out clustering independently in these feature spaces. As a result, we can obtain a certain number of clusters from each feature space, which paves the way for implementing our cross-reference strategy. In our scheme, N Cuts clustering algorithm, one of the popular spectral clustering algorithms, is employed for clustering.

3. MULTIMODAL RERANKING SCHEME

To grasp what is embedded in a video, human hearing is another necessary receptor apart from human vision, i.e., the video itself is generally endowed with multiple information sources. Hence, fusing information from multiple modalities, i.e., multimodal fusion for short, is a popular way currently to enhance the understanding of video content, which thereby helps to develop excellent video search engines. Likewise, video search re ranking can also benefit from multimodal fusion, especially when the size of the returned result set is relatively small. Based on the idea, a multimodal re-ranking scheme called CR-Re ranking is proposed.

3.1 Overview

The framework of CR-Re ranking is illustrated in denotes the initial result list ranked according to text-based search scores. The initial result list is processed individually in two distinct feature spaces, i.e., feature spaces A and B. In each feature space, all the results are first clustered into three clusters, and then the resulting clusters are mapped to three predefined rank levels, High, Median, and Low, in terms of their relevance to the query. Finally, a unique and improved shot ranking is formed by hierarchically combining all the ranked clusters from two different spaces. Note that only two modalities (or features) are considered here; however, the system can be easily extended to more modalities (or features).

3.2 Multi space Clustering

As illustrated in, we handle the initial search results by performing clustering and cluster ranking operations separately in two feature spaces. Clustering the initial search

1. Average numbers of the relevant shots at different depths. One group of bins corresponds to the case of 76 runs while the other corresponds to our run (BJTU). Note that, for the case of 76 runs, each average number is further averaged over 76 runs. Participants are required to return a ranking of 1,000 shots for each query topic and to submit at least one run (including 24 rankings where one ranking corresponds to one topic) for performance evaluation. In TRECVID'06, 76 runs, which are obtained mainly from text-based video search engines, are submitted, including the run (named as BJTU) from our developed video search system. Analyzing the retrieval effectiveness of these runs, we can reveal the weakness of current video search engines. Here, the average numbers of the relevant shots at different depths of the result list are used as the evaluation criterion. Given a depth X, the average number at depth X can be obtained by averaging the numbers of relevant shots in the top-X results over all 24 rankings.

3.3 Ranking at the Cluster Level

After several clusters are obtained from one feature space, the next step in our scheme is to indelicately rank them by their relevance to the query. To this end, some query-relevant shots should be selected in advance to convey the query intent. Similar to our selecting approach is also inspired by the PRF method. That is, the top-ranked initial results are considered as the informative shots. Here, the top-30 results are selected. Compared with directly treating these shots as relevant shots or adopting "soft" pseudo labels strategy , the proposed scheme only chooses K most informative shots from them by exploiting the centralization and decentralization properties. By doing this, some irrelevant shots (i.e., noisy points) can be filtered out effectively. Specifically, let A ¼ fa$_1$; a$_2$; . . . ; a$_{30}$g be the set of the top-30 shots.

As we have analyzed in Section 2, the relevant results in the top-30 shots usually group together in visual feature space, yet the irrelevant shots are scattered. It means that the distances between relevant shots are smaller than those distances between irrelevant shots or between relevant shots and irrelevant shots. Therefore, K shots with the smallest md distances are more possible to be the shots conveying the query intent, which can be selected to form the query-relevant shot set E. The value of K is selected empirically and fixed to 10. As mentioned previously, low-level features are more suitable for discriminating different shots within a finite shot set. In our case, the initial result list of 1,000 shots used for re ranking is a relatively small shot set. Hence, it is possible to nicely partition the initial list into several clusters in certain low-level feature spaces. Specifically, after extracting multiple features for each shot, we carry out clustering independently in these feature spaces. As a result, we can obtain a certain number of clusters from each feature space, which paves the way for implementing our cross-reference strategy. In our scheme, N Cuts clustering algorithm one of the popular spectral clustering algorithms, is employed for clustering.

irrelevant shots. Therefore, K shots with the smallest md the irrelevant shots are scattered. It means that the distances distances are more possible to be the shots conveying the query intent, which can be selected to form the query-relevant shot we have analyzed in Section 2, the relevant results in the top-30 shots usually group together in visual feature space, yet

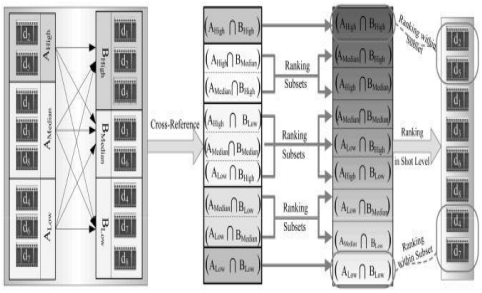

Fig. 4. Illustration of proposed fusion strategy.

where N is the number of clusters, and $A_i \setminus B_j$ stands for the intersection of clusters A_i and B_j.

As a matter of fact, the rank levels of subsets cannot be compared using merely the above criteria if (i þ j) is equal to (m þ n), just like the intersections $(A_1 \setminus B_2)$ and $(A_2 \setminus B_1)$.

So far, an ordered subset list has been formed. Although the ranks of shots in different subsets can be compared by the ranks of their corresponding subsets, we do not know which shot within the same subset is more relevant to the query. Hence, we need to find a method to order the shots within the same subset, i.e., ranking at the shot level. Here, the score or rank information of the initial ranking is used to order these shots.

4 EXPERIMENTS

4.1 Data Set and Evaluation Criteria

We experimentally validate our re ranking scheme on the NIST TRECVID'06 benchmark data set. The data set consists of approximately 343 hours of MPEG-1 broadcast news videos, which is divided into 169 hours of development videos and 174 hours of test videos. In all experiments, only the test data set is used for evaluation.

In video search scenarios, a shot is treated as the fundamental unit. Therefore, feature extraction is based on shots. For each video shot, four modalities are extracted: 78-D textual feature (TEXT) constructed from ASR/MT transcripts 120-D visual feature (MM) used in 73-D edged direction histogram (EDH), and 225-D grid color moment (GCM) For the performance evaluation, TRECVID suggests a number of criteria [28]. Three of them are employed in our evaluation, including precision at different depths of result list (Prec_D), non interpolated average precision (AP), and mean average precision (MAP). We denote D as the depth where precision is computed. Let S be the total number of returned show here T_n is the nth query topic, F_i ¼ 1 if the ith shot is relevant to the query and 0 otherwise, R stands for the total number of true relevant shots, and N denotes the number of query topics. Prec_D is utilized to assess the precision at different depths of the result list. AP shows the performance of a single query topic, which is sensitive to the entire ranking of documents. MAP summarizes the overall performance of a search system over all the query topics. Note that only the top-100 shots in the re ranked result list are considered for computing both AP and MAP.

4.2 Text-Only Baseline

The basic idea of text-based video search approach is to convert video retrieval into text

document search. Given a query text by users, the system then returns a series of approximately relevant video shots by matching the query text with the text documents associated with the video shots. In our prior work [11], we have constructed a fully automatic text-based video search engine exploring speech transcripts. Using this search engine, we can obtain an initial search list of 1,000 shots for each query topic.

TABLE 1

Comparison of Different Cluster Numbers

System	MAP	Gain
Text-only baseline	0.0333	0.0%
NCut+3	0.0454	36.3%
NCut+5	0.0443	33%
NCut+10	0.0378	13.5%
NCut+15	0.0328	-1.5%
NCut+30	0.0262	-21.3%

Reordering the initial list, our proposed re ranking scheme leads to high accuracy on the top-ranked results.

4.3 Number of Clusters

In our case, the number of groups is identical to the number of rank levels used in cluster ranking stage. Generally, varying cluster number should not have a significant impact on the re ranking performance, as stated. However, the performance of proposed method is sensitive to the number of clusters due to the limitation of cluster ranking. As stated in Section 3.3, the clusters can only be coarsely ranked according to their similarity to a noisy query-relevant shot set E. If the initial results are partitioned into too many clusters (or rank levels), the effect of noise will significantly violate the correctness of cluster ranking, which thereby deteriorates the re ranking performance.

We have performed experiments to evaluate the sensibility of performance to the cluster number. In these experiments, the number of clusters varies from small to large, whereas the feature combination keeps unchanged. Here, only TEXT feature and MM feature are used.

The experimental results are shown in Table 1. As expected, increasing the number of clusters leads to worse performance, and the search quality is even worse than the text-only baseline when the cluster number is greater than 15. In the following experiments, the number of clusters is fixed to 3 unless noted otherwise.

4.4 Evaluation on Different Re ranking Methods

The proposed scheme is compared with several available methods for video search re ranking in this section. All these re ranking methods are conducted using only the TEXT feature and MM visual feature, which are constructed as follows:

Single-Re ranking. This kind of re ranking method is constructed by performing clustering and cluster ranking once in only one modality space. Here, two systems are built individually in the WRITING and MM feature spaces, namely, Single-TEXT and Single-MM.

Early-Fusion Re ranking: We construct this scheme by clustering and cluster ranking once in a single feature space. The main difference from Single-Re ranking is that, instead of using only one modality, the feature vector used in Early-Fusion is formed by concatenating the vectors of multiple modalities. Here, we only concatenate the TEXT feature vector and MM visual feature vector.

Late-Fusion Re ranking. The clustering results from two feature spaces (i.e., TEXT and MM spaces) are directly fused by randomly intersecting any two clusters from different modalities and then ranking the newly formed subset list. Compared with the proposed method, the Late-Fusion scheme skips the cluster ranking step before combination.

Table 2 summarizes the evaluation results of different methods. We should note that all the re ranking schemes clearly outperform the text-only baseline. It means the re ranking is indeed an effective manner for improving the search quality. Compared with other re ranking methods, C R-Re ranking achieves higher accuracy on the top-ranked shots. As shown in Table 2, although CR-Re ranking does not achieve the best overall performance (MAP), it gives performance that is more outstanding at all depths within the top-30 results. That is, CR-Re ranking pays much more attention to the precision improvement on the top-ranked results. From the perspective of multimodal fusion, while the overall performance of Early-Fusion (0.0451) is roughly as good as CR-Re ranking (0.0454), its Prec_D values within the top-30 results are far lower than the proposed method.

This clearly exhibits the advantage of cross-reference-based

TABLE -2

Comparison of Different Re ranking Methods

System	MAP(gain)	Prec_5	Prec_10	Prec_15	Prec_20	Prec_30	Prec_100
Text-only baseline	0.0333(0%)	0.1167	0.1375	0.1222	0.125	0.1264	0.0987
Single-TEXT	0.0398(19.5%)	0.1583	0.1708	0.1472	0.1375	0.1347	0.0908
Single-MM	0.0461(38.4%)	0.1750	0.1792	0.1611	0.1437	0.1361	0.1062
Early-Fusion	0.0451(35.4%)	0.1500	0.1667	0.1583	0.1458	0.1347	0.1042
Late-Fusion	0.0412(23.7%)	0.1417	0.1542	0.1611	0.1479	0.1389	0.1096
CR-Reranking	0.0454(36.3%)	0.2167	0.2042	0.1889	0.1646	0.1486	0.0992

TABLE-3

Performance Evaluation on Different Single-Re ranking Methods

System	MAP(gain)	Prec_5	Prec_10	Prec_15	Prec_20	Prec_30	Prec_100
Text-only baseline	0.0333(0%)	0.1167	0.1375	0.1222	0.125	0.1264	0.0987
Single-TEXT	0.0398(19.5%)	0.1583	0.1708	0.1472	0.1375	0.1347	0.0908
Single-MM	0.0461(38.4%)	0.1750	0.1792	0.1611	0.1437	0.1361	0.1062
Single-GCM	0.0489(46.8.7%)	0.15	0.1708	0.1611	0.1458	0.1458	0.1087
Single-EDH	0.0376(12.9%)	0.1167	0.1292	0.1333	0.1312	0.1236	0.1029

on the whole search quality by 23.7 percent, the proposed method performs better than it, especially the precision improvement on the top-ranked results. Indeed, it is reason-able to achieve this result, as cluster ranking method used in our work is a bit sensitive to the number of rank levels (or the number of subsets). After combining all the clusters from two feature spaces by intersection, the Late-Fusion scheme generates more subsets that are unfavorable for correctly determining their ranks.

4.5 Evaluation on Feature Combination

As stated in Section 3.1, multi view strategy is based on the assumption that features from different views should be compatible and uncorrelated with each other. In this section, we have performed experiments to evaluate the performance sensitivity to the settings of various feature combinations. In our experiments, all the extracted four feature types are used for evaluation. In addition to Single-TEXT and Single-MM, we also conduct additional two Single-Re ranking methods separately in GCM and EDH feature spaces, namely Single-GCM and Single-EDH. In Table 3, the evaluation results on the four Single-Re ranking methods are displayed, which indicate the effectiveness of the corresponding features. The performance of the proposed re ranking method is evaluated with varying feature combinations (i.e., TEXT+MM, TEXT+GCM, TEXT +EDH, GCM+MM, GCM+EDH, and MM+EDH). Table 4 summarizes the experimental results.

As seen in Table 4, when the TEXT feature combines with different visual features, only TEXT+EDH scheme achieves better overall performance than any of its corresponding Single-Re ranking schemes (i.e., Single-TEXT and Single-EDH), 0.0405 compared to 0.0398 and 0.0376. The reason is that the effectiveness of TEXT feature (0.0398) is more compatible with EDH feature (0.0376) than with any of the other two visual features. That is, the cross-reference-based multimodal fusion is indeed sensitive to the incompatibility of features. For evaluating the effect of uncorrelated assumption, we perform the experiments by combining different visual features (i.e., GCM+MM, GCM+EDH, and MM+EDH). Intuitively, these visual features are correlated with each other. From experimental results, however, we cannot conclude that the uncorrelated assumption has a strong impact on the cross-reference-based multimodal fusion, as GCM+MM scheme also performs better than any of its corresponding Single-Re ranking schemes (i.e., Single-GCM and Single-MM), 0.0506 compared to 0.0489 and 0.0461. As one possible explanation, the performance improvement of GCM+MM is due to the compatibility of GCM and MM features. In brief, compared with the uncorrelated assumption, the incompatibility has stronger impact on the cross-reference-based multimodal fusion.

Considering the evaluation results in Tables 3 and 4, almost all the cross-reference-based fusion schemes outer-form their corresponding Single-Re ranking schemes on the top-ranked results. For example, even though TEXT+GCM scheme results

in a smaller MAP than Single-GCM scheme, 0.0422 compared to 0.0489, it achieves more outstanding performance in precision at all the depths within the top-30 results. It shows the merit of the cross-reference-based multimodal fusion again.

4.6 Performance Analysis on all Query Topics

In this section, we evaluate our proposed preparation on varied query topics. Fig. 5 illustrates the statistics on APs across 24 query topics used in TRECVID'06 evaluation. The results show that the proposed re ranking scheme works well for named persons and named objects, such as "D. Cheney" and "Boats," as the search quality on these topics can benefit from the TEXT feature used in our scheme. More-over, our approach is also suitable for some query topics that are of distinctive visual properties, such as "soccer goalposts" and "scenes with snow." Similarly, prominent improvement is due to the use of the MM visual feature.

However, the search performance after re ranking is even below the performance of text-only baseline for some topics with motion properties, like "leaving a vehicle." The reason is that features used in our scheme lack the capability to capture motion properties in video. Hence, new research fruits in precise representation of shot will provide much more room for performance improvement. In addition, our proposed method also fails in some query topics with very few relevant shots within the top-30 results, such as "meeting" and "people with uniform." It is because cluster ranking is based essentially on the relevant shots within the top-30 results. Incorrectly ranked clusters will deteriorate the re ranking performance.

REFERENCES

[1] M.S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-Based Multimedia Information Retrieval: State of the Art and Chal-lenges," ACM Trans. Multimedia Computing, Comm., and Applica-tions, vol. 2, pp. 1-19, 2006.

[2] A. Smeaton and T. Ianeva, "TRECVID-2006 Search Task," TREC Video Retrieval Evaluation Online Proc., 2006.

[3] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," Proc. ACM SIGKDD, pp. 133-142, 2002.

[4] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, "Analysis of a Very Large Web Search Engine Query Log," ACM SIGIR Forum, vol. 33, pp. 6-12, 1999.

[5] Y. Cao, J. Xu, T.-Y. Liu, H. Li, Y. Huang, and H.-W. Hon, "Adapting Ranking SVM to Document Retrieval," Proc. 29th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, 2006.

[6] C.G.M. Snoek, J.C. van Gemert, J.M. Geusebroek, B. Huurnink, D.C. Koelma, G.P. Nguyen, O. de Rooij, F.J. Seinstra, A.W.M. Smeulders, C.J. Veenman, and M. Worring, "The MediaMill TRECVID 2005 Semantic Video Search Engine," TREC Video Retrieval Evaluation Online Proc., 2005.

[7] A. Amir, J. Argillander, M. Campbell, A. Haubold, G. Iyengar, S. Ebadollahi, F. Kang, M.R. Naphade, A. Natsev, J.R. Smith, J. Tesic, and T. Volkmer, "IBM Research TRECVID-2005 Video Retrieval System," TREC Video Retrieval Evaluation Online Proc., 20