ISSN:0975-9646

R.Ranjani et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7690-7696

An Enhancement of Mining Videos Based On Annotation Reranking Method Using Ncut Algorithms

R.Ranjani[1]. S. Selva Seeba [2]

[1]*Sri Krishna Arts and Science College,India*
[2]*PSG College Of Arts and Science,India*

Abstract-Annotation Reranking, this combines multimodal features in the manner of cross reference. The fundamental idea of Annotation Reranking 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. Multiview 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. Annotation Reranking 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. In our scheme, NCuts clustering algorithm is employed for clustering.

Key words:-Content-based video search, NCUT algorithm, annotation reranking, cluster, multiview learning.

1. INTRODUCTION

As an emerging research field, content-based video retrieval (CBVR) has attracted a great deal of attention in recent years. While various retrieval models have been developed to improve 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 human 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.

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. 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 reranking 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. Analysis on click- through data from a very large Web search engine log also reflects such preference. Therefore, it is more crucial to offer high accuracy on the top-ranked documents than to improve the whole search performance on the entire result list.

Many methods have been proposed 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, RF-based methods require user's 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. For instance, the coretrieval algorithm 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 Adaboost-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 metasearch strategy which is originally put forward in the field of information retrieval is imported to CBVR for improving video retrieval effectiveness. The key idea of metasearch is that multiple result lists returned by several different search engines in response to a given query are aggregated into a single list in an optimal way. Metasearch is generally based on the "unequal 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 Page Rank like graph based approach and the model-based reranking algorithm.

As a kind of multimodal fusion method, metasearch can simultaneously leverage multiple ranked lists from several search engines based on various modalities. However, a general problem with metasearch is that it is usually hard to expect users to provide query examples with multimodal representations. In addition, it is not easy in practice to get access to multiple search engines based on different modalities.

As an alternative scheme, the reranking method can improve search quality by reordering the initial result list. Although the total number of relevant documents remains fixed after reranking, the precision improvement at the low depth of the result list can be expected by forcing true relevant documents to move forward. Traditionally, this kind of technique is used in the field of Web search. The predominant work includes Page Rank and HITS In the multimedia search community, the idea of reranking has been extended to develop advanced video search engines. As a successful attempt, IB-Reranking, based on the Information Bottleneck (IB) principle, explores multimodal cues to reorder the initial search results. 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 Reranking 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 etal'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.

Annotation Reranking, this combines multimodal features in the manner of cross reference. The fundamental idea of Annotation Reranking 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. Multiview 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. Multiview 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 Annotation Reranking 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.

1.1 ANNOTATION STRATEGY DIFFERS IN TWO WAYS FROM METASEARCH

1. 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.
2. The second one is that, instead of using multiple lists at the shot level, we first coarsely rank each list at the cluster level, and then integrate all the resulting clusters hierarchically,Also achieves higher accuracy on the top-ranked shots.

2. PROBLEM SPECIFICATION

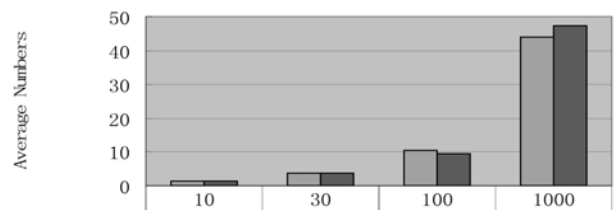
Most of video retrieval implement search procedure by implicitly or explicitly measuring the similarity between the query and database shots in some low-level feature spaces. However, such similarity is not usually consistent with human perception due to the limitation of current image/video understanding 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.

2.1. PROBLEM ANALYSIS

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

2.2 WEAKNESS OF CURRENT SEARCH ENGINES

As a well-recognized community for video search, NIST TRECVID [28] provides 24 query topics for all participants to test their video search systems.



	10	30	100	1000
76 runs	1.41	3.83	10.55	43.96
BJTU	1.38	3.79	9.5	47.33

Depths

Fig.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.

In annual competition all 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 rankings. In Fig. 1, we illustrate the statistical results on both the BJTU run and all the 76 runs.

From Fig.1, it is not hard to reach a conclusion that relevant shots are scarce in the top-ranked results, e.g., 1.41 for depth 10 and 3.83 for depth 30. However, in real-world application scenarios, users merely restrict their attention to the top-ranked shots. That is the current video search engines cannot satisfy user information needs. Hence, it is of great importance to develop some new methods that achieve higher accuracy on the top-ranked shots.

2.3 FEASIBILITY FOR ALLEVIATING THE WEAKNESS

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 =1000, average number = 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. In addition, some observations on the initial rankings are helpful in building an effective reranking 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.

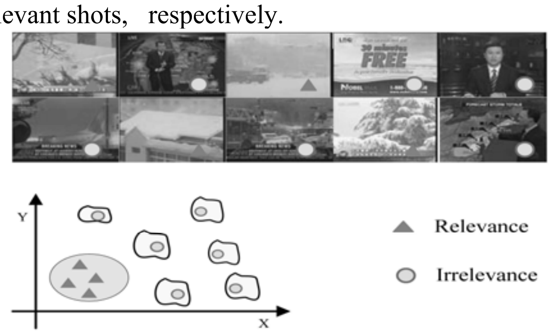


Fig.2 Diagrammatic sketch of centralization and decentralization properties in a 2D visual feature space.

Fig. 2 demonstrates both properties in a 2D visual feature space, where four gathered relevant shots (triangles) indicate the centralization attribute while six scattered irrelevant shots (circles) display the decentralization property. Based on these observations, we can give a fairly effective solution for selecting some query-relevant shots, which are important for cluster ranking in our approach. 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 reranking procedure.

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 reranking can also benefit from multimodal fusion, especially when the size of the returned result set is relatively small. Based on the idea, a multimodal reranking scheme called Annotation-Reranking is proposed.

3. 1. OVERVIEW

The framework of Annotation-Reranking is illustrated in Fig. 3, where {d1, d2..., d8} 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, i.e., 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. THE NCUT ALGORITHM

Let:

$$d(i) = \sum_j w_{ij}$$

Also, let D be an $n \times n$ diagonal matrix with *d* on the diagonal, and let *W* be an $n \times n$ symmetrical matrix with $W_{ij} = w_{ij}$.

After some algebraic manipulations, we get:

$$\min_{(S,\overline{S})} ncut(S, \overline{S}) = \min_{y} \frac{y^T(D-W)y}{y^T D y}$$

Subject to the constraints:

$y_i \in \{1, -b\}$, for some constant $-b$

$y^t D 1 = 0$

$$\frac{y^T(D-W)y}{y^T D y}$$

Minimizing subject to the constraints above is NP-hard. To make the problem tractable, we relax the constraints on *y*, and allow it to take real values. The relaxed problem can be solved by solving the generalized eigenvalue problem $(D - W)\, y = \lambda D y$ for the second smallest generalized eigenvalue.

3.3. MULTISPACE CLUSTERING

As illustrated in Fig. 3, we handle the initial search results by performing clustering and cluster ranking operations separately in two feature spaces. Clustering the initial search results, we can obtain three clusters from each feature space, which are needed for the hierarchical fusion in the following steps.

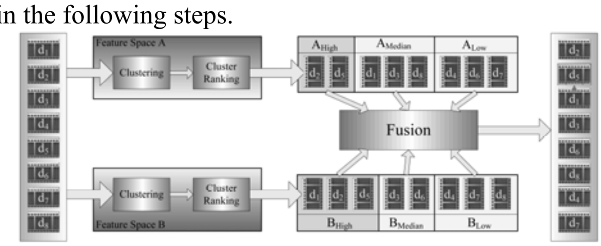


Fig.3: Framework of proposed Annotation-Reranking method.

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 reranking 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, NCUTs clustering algorithm, one of the popular spectral clustering algorithms, is employed for clustering.

Ranking at the Cluster Level After several clusters are obtained from one feature space, the next step in our scheme is to coarsely 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= {a1, a2. . . ,a3} be the set of the top-30 shots. They are ranked in ascending order according to the following distance:

$$md(a_i, A\backslash a_i) = \min_{a_j \in A\backslash a_i} \{d(a_i, a_j)\}, \quad (1)$$

Where d(.,.) is the Euclidean distance, 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.Therefore; the implementation of cluster ranking is equivalent to measuring the similarity between the set E and the clusters. For measuring the relevance between shot sets, we employ the modified Hausdorff distance, which is defined as follows:

$$hd(E, C) = \underset{e \in E}{mean}\left\{\min_{c \in C}\{d(e, c)\}\right\}, \quad (2)$$

Where E is the query-relevant set and C can be a cluster or any shot set. Note that hd(E, C) is a directed Hausdorff distance from E to C. Following (2), we can assign corresponding ranks to the clusters in each modality space.

3.4. ANNOTATION-BASED FUSION STRATEGY

Our final goal is to obtain a unique and improved reranking of the initial results, especially paying more attention to the accuracy on the top-ranked shots. In order to move vigorously toward this goal, we hierarchically fuse all the ranked clusters from different modalities using a annotation strategy. Fig. 4 illustrates the schematic diagram of our fusion method with three rank levels (i.e., High, Median, and Low). As shown in Fig. 4, our fusion approach is composed of three main components: combining these ranked clusters using cross-reference strategy, ranking subsets with the same rank level, and ranking shots within the same subset. Note that the rank levels are denoted numerically in the following formulas for the convenience of expression. The rank levels High, Median, and Low in Fig. 4 are equivalent to the rank levels 1, 2, and 3, respectively. We assume that a shot has a high rank if it exists simultaneously in multiple high-ranked clusters from different modalities. Based on this assumption, we put forward a cross-reference strategy to hierarchically combine all the ranked clusters, leading to a coarsely ranked subset list. Specifically, let $\{A_1, A_2, . . ., A_N\}$ and $\{B_1, B_2, . . ., B_N\}$ be the sets of the ranked clusters from feature spaces A and B, respectively, and Rank be the operation of measuring the rank level of a cluster or shot. The ranked clusters in each set are arranged from high-rank level to low-rank level in ascending order of their subscripts, that is, Rank(Ai) is greater than Rank(Aiþ1). Then, two ranked cluster sets can be integrated into a unique and coarsely ranked subset list according to the following inference rule:

$$\begin{aligned} &Rank(A_i \cap B_j) \succ Rank(A_m \cap B_n), \\ &\quad if\ (i+j) \prec (m+n), i, j, m, n = 1, \ldots, N, \end{aligned} \quad (3)$$

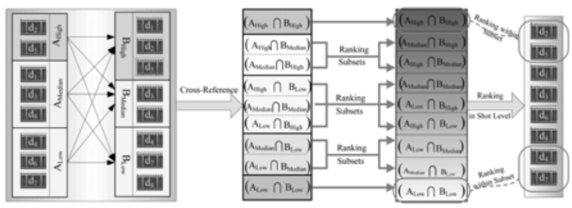


Fig.4 proposed Annotation-Reranking method.

Where N is the number of clusters, and $A_i \cap B_j$ stands for the intersection of clusters Ai and Bj.As a matter of fact, the rank levels of subsets cannot be compared using merely the above criteria if (i + j) is equal to (m + n), just like the intersections (A1 ∩ B2) and (A2 ∩ B1). To address this issue, we employ the method used in the cluster ranking step to order those subsets, which can be formulized as follows:

$$\begin{aligned} &Rank(A_i \cap B_j) \succ Rank(A_m \cap B_n), \\ &\quad if\ (i+j) = (m+n), hd(E, A_i \cap B_j) \prec hd(E, A_m \cap B_n), \end{aligned} \quad (4)$$

Where the distance hd(.,.)can be computed in any of the feature spaces. 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. The ranking rule is defined as follows:

$$Rank(d_m) \succ Rand(d_n),\ if\ S_m \succ S_n, \qquad (5)$$

Where dm and dn denote shots m and n within the same subset, respectively, Sm and Sn correspond to the scores or ranks of shots m and n, respectively.

3.5. NUMBER OF CLUSTERS

The number of clusters 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 reranking performance. The performance of proposed method is sensitive to the number of clusters due to the limitation of cluster ranking. 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 reranking performance. Performed experiments to evaluate the sensibility of performance to the cluster number. The number of clusters varies from small to large, whereas the feature combination keeps unchanged. Here, only TEXT feature and MM feature are used. 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.

4. RESULTS AND DISCUSSIONS

All reranking methods are conducted using only the TEXT feature and MM visual feature, which are constructed as follows:

1. Single-Reranking: This kind of reranking method is constructed by performing clustering and cluster ranking once in only one modality space. Here, two systems are built individually in the TEXT and MM feature spaces, namely, Single-TEXT and Single-MM.
2. Early-Fusion Reranking: Construct this scheme by clustering and cluster ranking once in a single feature space. The main difference from Single-Reranking 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.
3. Late-Fusion Reranking: 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.

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 shots and Ri the number of true relevant shots in the top-i returned results. Then, these evaluation criteria can be defined as follows:

$$\Pr ec\_D(T_n) = \frac{1}{D}\sum_{i=1}^{D} F_i, \qquad (6)$$

$$AP(T_n) = \frac{1}{R}\sum_{i=1}^{S}\left(\frac{R_i}{i}\cdot F_i\right), \qquad (7)$$

$$MAP = \frac{1}{N}\sum_{n=1}^{N} AP(T_n), \qquad (8)$$

where Tn is the nth query topic, Fi = 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 reranked result list are considered for computing both AP and MAP.

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.1 Comparison of Different Reranking Methods

Table 1 summarizes the evaluation results of different methods. All the reranking schemes clearly outperform the text-only baseline. It means that reranking is indeed an effective manner for improving the search quality. Compared with other reranking methods, Annotation-Reranking achieves higher accuracy on the top-ranked shots. As shown in Table 3, although Annotation-Reranking 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-Reranking 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 Annotation -Reranking (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 multimodal fusion. Particular mention should be made to the performance comparison between fusion-based reranking methods and Single-Reranking methods. Table 3 also shows that the Single-MM reranking results in a larger improvement on MAP than any of the three fusion-based reranking methods (i.e., Early-Fusion, Late-Fusion, and Annotation-Reranking), 38.4 percent compared to 35.4, 23.7, and 36.3 percent. The reason is that the performance of multimodal fusion is under the constraint of compatibility, which will be discussed in more detail in the next section. Even so, Annotation-Reranking still performs better on the top-ranked results than Single-MM reranking.

Comparing Late-Fusion with CR-Reranking, although the Late-Fusion scheme also achieves a significant improvement 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 reasonable 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 inter section, the Late-Fusion scheme generates more subsets that are unfavorable for correctly determining their ranks.

Annotation –reranking 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 reranking 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, Annotation –reranking 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 reranking performance.

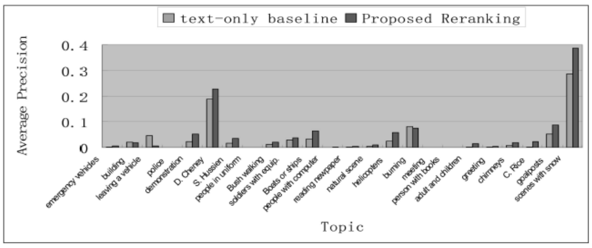


Fig.5 Performance of the proposed reranking scheme

Fig. 5 illustrates the statistics on APs used in TRECVID’06 evaluation. The results show that the proposed reranking scheme works well for named persons and named objects, as the search quality on these topics can benefit from the TEXT feature used in our scheme. Moreover, our approach is also suitable for some query topics that are of distinctive visual properties. The search performance after reranking is even below the performance of text-only baseline for some topics with motion properties. 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. Cluster ranking is based essentially on the relevant shots within the top-30 results. Incorrectly ranked clusters will deteriorate the reranking performance.

5. Conclusion And Scope For Future Work

Annotation reranking method combines multimodal features via a cross-reference strategy. It can handle the initial search results independently in various modality spaces. Specifically, the initial search results are first divided into several clusters individually in different feature spaces. Then, the clusters from each space are mapped to the predefined ranks according to their relevance to the query. Given the ranked clusters from all the feature spaces, the cross-reference strategy can hierarchically fuse them into a unique and improved result ranking. Experimental results show that the search effectiveness, especially on the top ranked results, is improved significantly.

The proposed reranking method is sensitive to the number of clusters due to the limitation of cluster ranking. In the future, it is possible to develop a new method to adaptively choose cluster number for different feature spaces. In addition, new strategies are to be investigated for selecting query-relevant shots, e.g., using pseudo negative samples to exclude irrelevant shots.

References:

[1] M.S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-Based Multimedia Information Retrieval: State of the Art and Challenges," *ACM Trans. Multimedia Computing, Comm., and Applications,* 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. Tešić, and T. Volkmer, "IBM Research TRECVID-2005 Video Retrieval System," *TREC Video Retrieval Evaluation Online Proc.,* 2005.

[8] S.F. Chang, W.H. Hsu, L. Kennedy, L. Xie, A. Yanagawa, E. Zavesky, and D.-Q. Zhang, "Columbia University TRECVID-2005Video Search and High-Level Feature Extraction," *TREC Video Retrieval Evaluation Online Proc.,* 2005.

[9] A.G. Hauptmann, M. Christel, R. Concescu, J. Gao, Q. Jin, W.-H. Lin, J.-Y. Pan, S.M. Stevens, R. Yan, J. Yang, and Y. Zhang, "CMU Informedia's TRECVID 2005 Skirmishes," *TREC Video Retrieval Evaluation Online Proc.,* 2005.

[10] J.H. Yuan, W.J. Zheng, L. Chen, D.Y. Ding, D. Wang, Z.J. Tong, H.Y. Wang, J. Wu, J.M. Lin, and B. Zhang, "Tsinghua University at TRECVID 2005," *TREC Video Retrieval Evaluation Online Proc.,* 2005.

[11] S.K. Wei, Y. Zhao, Z.F. Zhu, N. Liu, Y.F. Zhao, L. Zhang, and F. Wang, "BJTU TRECVID 2006 Video Retrieval System," *TREC Video Retrieval Evaluation Online Proc.,* 2006.

[12] W.H. Hsu, L.S. Kennedy, and S.-F. Chang, "Reranking Methods for Visual Search," *IEEE Trans. Multimedia,* vol. 14, no. 3, pp. 14-22, July-Sept. 2007.

[13] R. Yan and A.G. Hauptmann, "Co-Retrieval: A Boosted Reranking Approach for Video Retrieval," *IEE Proc. Vision, Image and Signal Processing,* vol. 152, pp. 888-895, 2005.

[14] J.H. Lee, "Analyses of Multiple Evidence Combination," *ACM SIGIR Forum,* vol. 31, pp. 267-276, 1997
[15] J.A. Aslam and M. Montague, "Models for Metasearch," *Proc. 24th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval,* pp. 276-284, 2001.
[16] J. Liu, W. Lai, X.S. Hua, Y. Huang, and S. Li, "Video Search Re-Ranking via Multi-Graph Propagation," *Proc. 15th Int'l Conf. Multimedia,*pp. 208-217, 2007.
[17] A. Haubold, A. Natsev, and M.R. Naphade, "Semantic Multimedia Retrieval Using Lexical Query Expansion and Model-Based Reranking,"*Proc. IEEE Int'l Conf. Multimedia and Expo,* 2006.
[18] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," Stanford Digital Libraries Working Paper, 1998.
[19] T.H. Haveliwala, "Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search," *IEEE Trans. Knowledge and Data Eng.,* vol. 15, no. 4, pp. 784-796, July/Aug. 2003.
[20] J.M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," *J. ACM,* vol. 46, pp. 604-632, 1999.
[21] W.H. Hsu, L.S. Kennedy, and S.-F. Chang, "Video Search Reranking via Information Bottleneck Principle," *Proc. 14th Ann. Int'l Conf. Multimedia,* pp. 35-44, 2006.
[22] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury, "How Flickr Helps Us Make Sense of the World: Context and Content in Community-Contributed Media Collections," *Proc. 15th Int'l Conf. Multimedia,* pp. 631-640, 2007.
[23] A. Blum and T. Mitchell, "Combining Labeled and Unlabeled Data with Co-Training," *Proc. 11th Ann. Conf. Computational Learning Theory,* pp. 92-100, 1998.
[24] K. Nigam and R. Ghani, "Understanding the Behavior of Co-Training," *Proc. KDD-2000 Workshop Text Mining,* 2000.
[25] Z.H. Zhou and M. Li, "Tri-Training: Exploiting Unlabeled Data Using Three Classifiers," *IEEE Trans. Knowledge and Data Eng.,* vol. 17, no. 11, pp. 1529-1541, Nov. 2005.
[26] I. Muslea, S. Minton, and C. Knoblock, "Active+ Semi-Supervised Learning $=$ Robust Multi-View Learning," *Proc. Int'l Conf. Machine Learning,* pp. 435-442, 2002.
[27] R. Yan and M. Naphade, "Multi-Modal Video Concept Extraction Using Co-Training," *Proc. IEEE Int'l Conf. Multimedia and Expo,* pp. 514-517, 2005.
[28] TRECVID, "TREC Video Retrieval Evaluation," http:// www-nlpir.nist.gov/ projectstrecvid/, 2009.
[29] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 888-905, Aug. 2000.
[30] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge, "Comparing Images Using the Hausdorff Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 9, pp. 850-863, Sept. 1993.
[31] A. Yanagawa, W.H. Hsu, and S.-F. Chang, "Brief Descriptions of Visual Features for Baseline TRECVID Concept Detectors," ADVENT Technical Report #219-2006-5, Columbia Univ., July 2006.