

A Survey on Visual Search Reranking

Thalla Shankar¹, Lalitha Manglaram², Murali Sadak³

¹(Assistant Professor, Department of CSE, Kshatriya College of Engineering, JNTUH, Andhra Pradesh, India)

²(Assistant Professor, Department of CSE, Kshatriya College of Engineering, JNTUH, Andhra Pradesh, India)

³(Assistant Professor, Department of CSE, Kshatriya College of Engineering, JNTUH, Andhra Pradesh, India)

Abstract : Search reranking is considered as a best and common way to improve retrieval precision. However the problem is not trivial especially when we are considering multiple features or modalities for search in image and video retrieval. This paper suggests a new kind of reranking algorithm, the circular reranking, that supports the mutual exchange of information across multiple modalities for improving search performance, and follows the philosophy of strong performing modality could learn from weaker ones, and weaker modality does benefit by interacting with stronger modalities. Technically, circular reranking conducts multiple cyclic runs of random walks by exchanging the ranking scores among different features. Contrasting the existing techniques, the reranking procedure reassures interaction between various modalities to seek consensus which are useful for reranking. In this paper, we study several circular reranking properties, including how and which order of information propagation should be configured to fully exploit the potential of modalities for reranking.

Keywords: Visual search, circular reranking, multimodality fusion.

I. Introduction

THE rapid development of Web 2.0 technologies has led to the surge of research activities in visual search. While visual documents are rich in audio-visual content and user-supplied texts, commercial visual search engines to date mostly perform retrieval by keyword matching. A common practice to improve search performance is to rerank the visual documents returned from a search engine using a larger and richer set of features. The ultimate goal is to seek consensus from various features for reordering the documents and boosting the retrieval precision. There are two general approaches along this direction: visual pattern mining [6] and multi-modality fusion [1], [2]. The former mines the recurrent patterns, either explicitly or implicitly, from initial search results and then moves up the ranks of visually similar documents. Random walk [7], for instance, performs self-reranking through identifying documents with similar patterns based on inter-image similarity and initial rank scores. This category of approaches, nevertheless, seldom explores the joint utilization of multiple modalities. Instead, every different modalities treated as independent modality. Furthermore, the utilization of a modality is often application dependent, making it difficult to generalize the mining for general-purpose search. Multi-modality fusion, in contrast, predicts the importance of modalities, for instance, through fusion weight learning, and linearly combines them for reordering documents. The fusion, however, is done at the decision stage. Mainly, the fusion weights estimation is obtained from the ranking scores in different ranked lists.

This paper suggests a different algorithm, named circular reranking, that takes advantages of both pattern mining and multi-modality fusion for visual search. Importantly, modality interaction is considered, on one hand to implicitly mine repeated patterns, and the other, to leverage the modalities of different strength for maximizing search performance.

A. Circular reranking

The basic idea of circular reranking is to facilitate interaction among different modalities through mutual reinforcement. In this way, the performance of strong modality is enhanced through communication with weaker ones, while the weak modality is also benefited by learning from strong modalities. The mutual exchange of information across multiple modalities for improving search performance, follows the philosophy of strong performing modality can learn from weaker ones, whereas weak modality can be benefited by interacting with stronger ones. Precisely, circular reranking conducts several randomized cyclic runs through exchanging the ranking scores among different features. Unlike the existing techniques, the reranking procedure encourages interaction between modalities to seek a harmony which is useful for reranking.

B. Multimodality fusion

In recent times, multimodal fusion has gained much attention of many researchers due to the benefit it provides for various multimedia analysis tasks. The combination of multiple media and their associated features, or the intermediate decisions in order to perform an analysis task is referred to as multimodal fusion. Modality interaction is considered on one hand to implicitly mine recurring patterns, and on the other, to leverage the modalities of different strength for maximizing search performance. A multimedia analysis task involves

processing of multimodal data in order to obtain valuable insights about the data, a situation, or a higher level activity. These media and related features are fused together for the accomplishment of various analysis tasks. The fusion of multiple modalities can provide complementary information and increase the accuracy of the overall decision making process. For example, fusion of audio-visual features along with other textual information have become more effective in detecting events from a team sports video, which would otherwise not be possible by using a single medium. The capturing and processing of media streams may involve certain costs, which may influence the fusion process. The cost may be incurred in units of time, money or other units of measure. For instance, the task of object localization could be accomplished cheaply by using a RFID sensor compared to using a video camera.

C. Visual search

Visual search is a type of perceptual task requiring attention that typically involves an active scan of the visual environment for a particular object or feature (the target) among other objects or features. Visual search can be done by either with or without eye movements. The ability to consciously locate an object (target) amongst a complex array of stimuli has been extensively studied over the past 40 years. Practical examples of this can be seen in everyday life such as picking out a on-the-shelf product in a supermarket shelf, animals looking for food amongst masses of leaves, trying to find your loved ones in a huge crowd and playing visual search tasks. Here is a variety of speculation about the origin and progress of visual search in humans. It is showing that the visual exploration of complex natural scenes, both humans and nonhumans make highly stereotyped eye movements.

II. Literature Survey

We briefly group the related works for visual search reranking into two categories: recurrent pattern mining and multimodality fusion. The former assumes the existence of common patterns among relevant documents for reranking. The later predicts or learns the contribution of a modality in search reranking.

A. Recurrent Pattern Mining

In this direction the research has carried on along three different dimensions: self-reranking [4], [7], [5], crowd-reranking by exploiting online crowd sourcing knowledge [10], and example-based reranking by leveraging user-provided queries [11], [20].

Fergus et al. [4], employed probabilistic Latent Semantic Analysis (pLSA) for mining visual categories through clustering of images in the initial ranked list and which extends pLSA (as applied to visual words) to include spatial information in a translation and scale invariant manner. Candidate images are then reranked based on the distance to the mined categories. Self-reranking seeks consensus from the initial ranked list as visual patterns for reranking.

Hsu et al. [5], employed information bottleneck (IB) reranking to find the clustering of images that preserves the maximal mutual information between the search relevance and visual features. The IB reranking method, based on a rigorous Information Bottleneck (IB) principle which finds the optimal image clustering that preserves the maximal mutual information between the search relevance and the high-dimensional low-level visual features of the images in the text search results. Among all the possible clustering's of the objects into a fixed number of clusters, the optimal clustering is the one that minimizes the loss of mutual information (MI) between the features and the auxiliary labels.

Richter et al. [12], employed an crowd-reranking is similar to self-reranking except that consensus is sought simultaneously from multiple ranked lists obtained from Internet resources and further formulated the problem as random walk over a context graph built through linearly fusing multi-modalities for visual search. We proposed to use a multimodal similarity measure to find nearest neighbors of images. The nearest neighbor search of an image is then limited to such a subspace, i.e. to a subset of images in our database. This way we reduce the number of image comparisons required for the graph construction to a linear amount depending on the cluster sizes.

Liu et al. [10], suggested a reranking paradigm by issuing query to multiple online search engines. Based on visual word representation, both concurrent and salient patterns are respectively mined to initialize a graph model for randomized walks based on reranking. Different from self- and crowd-reranking, example-based reranking relies on a few query examples provided by users for model learning.

Yan et al. [20], employed an classifiers are learnt by treating query examples as positive training samples while randomly picking pseudo-negative samples from the bottom of initial ranked list. The classifiers which capture the visual distribution of positive and negative samples are then exploited for reranking.

Liu et al. [11], proposed a query examples are utilized to identify relevant and irrelevant visual concepts, which are in turn employed to discover the rank relationship between any two documents using mutual information for correcting ranking of document pairs.

B. Multimodality Fusion

Multi-modality fusion based on weighted linear fusion is widely adopted. Broadly, we can categorize the existing research into adaptive [15], and query-class-dependent fusion [9].

Wilkins et al, [18], proposed a multi-modal data for video Information Retrieval, modeled the change of scores in a list to predict the importance of a modality. Specifically, the gradual (drastic) change of scores indicates the difficulty (capability) of a modality in distinguishing relevant from irrelevant items, and fusion weights are thus determined accordingly. Firstly that an examination of the distribution of the scores can reveal correlations between those results which undergo a rapid initial change in score, to those results which perform well with regard to relevance. Secondly, we presented an initial model to take advantage of these correlations and to automatically generate weights for a retrieval system without giving that system any prior training or outside knowledge of the collection.

Tan et al, [15], proposed an agreement-fusion optimization model for fusing multiple heterogeneous data. The leveraged rank agreement mined from multiple lists iteratively to update the weights of modalities until reaching an equilibrium stage. The agreement between the scores from multiple modalities is explored to guide the fusion of multiple graphs in both linear and adaptive manners. The agreement is exploited in two ways, namely as the personalization distribution for random walk, or as pseudo training samples for semi-supervised learning to adapt the fusion weights of different modalities. To reconcile the conflicting objectives between graph fusion and agreement, score exchange is conducted iteratively between the two steps to reach an equilibrium solution.

Kennedy et al, [9], proposed a query class dependent search models in multimodal retrieval for the automatic discovery of query classes. This scheme starts by predefining query classes, then learning of weights in offline conducted on the query class level. During search, a given query is routed into one of the predefined classes, and the learnt weights are directly applied for fusion. This scheme is effective in general when the underlying query classes can be clearly defined and there are enough samples for weight learning. Query-class-dependent models for multimodal search by defining query classes through a clustering process according to search method performance and semantic features.

Wei et al, [16], proposed a concept-driven multi-modality fusion (CDMF), explores a large set of predefined semantic concepts for computing multi-modality fusion weights in a novel way. In CDMF, the query-modality relationship is decomposed into two components that are much easier to compute: query-concept relatedness and concept-modality relevancy. In earlier, it can be efficiently estimated online by using semantic and visual mapping techniques, while the latter can be computed offline based on concept detection accuracy of each modality. To determine the fusion weights, the concept-to-modality relationship in a large number of visual concepts is mapped with the query. It automatically discovers useful query classes by clustering queries in a training set.

III. Conclusion

This paper presents a survey on various Circular Reranking algorithms that were proposed by earlier researches for the better development in the field of Image Processing. Various algorithms and methods discussed above will help in developing efficient and effective circular reranking for image processing. In the future scope, we will be presenting a comparative study of various algorithms for circular reranking.

References

- [1] P. K. Atrey, M. A. Hossain, A. E. Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: A survey," *Springer Multimedia Syst. J.*, vol. 16, no. 6, pp. 345–379, 2010.
- [2] C. A. Bhatt and M. S. Kankanhalli, "Multimedia data mining: State of the art and challenges," *Multimedia Tools Appl.*, vol. 51, no. 1, pp. 35–76, 2011.
- [3] K. M. Donald and A. F. Smeaton, "A comparison of score, rank and probability-based fusion methods for video shot retrieval," in *Proc. ACM Int. Conf. Image Video Retr.*, Jul. 2005, pp. 61–70.
- [4] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning object categories from google's image search," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 1816–1823.
- [5] W. Hsu, L. Kennedy, and S.-F. Chang, "Video search reranking via information bottleneck principle," in *Proc. ACM Int. Conf. Multimedia*, 2006, pp. 35–44.
- [6] W. Hsu, L. Kennedy, and S.-F. Chang, "Reranking methods for visual search," *IEEE Multimedia*, vol. 14, no. 3, pp. 14–22, Jul.–Sep. 2007.
- [7] W. Hsu, L. Kennedy, and S.-F. Chang, "Video search reranking through random walk over document-level context graph," in *Proc. ACM Int. Conf. Multimedia*, 2007, pp. 971–980.
- [8] L. Kennedy, S. F. Chang, and A. Natsev, "Query-adaptive fusion for multimodal search," *Proc. IEEE*, vol. 96, no. 4, pp. 567–588, Apr. 2008.
- [9] L. Kennedy, A. Natsev, and S. Chang, "Automatic discovery of queryclass-dependent models for multimodal search," in *Proc. ACM Int. Conf. Multimedia*, 2005, pp. 882–891.
- [10] Y. Liu, T. Mei, and X.-S. Hua, "CrowdReranking: Exploring multiple search engines for visual search reranking," in *Proc. ACM Special Interest Group Inf. Retr.*, 2009, pp. 500–507.

- [11] Y. Liu, T. Mei, X. Wu, and X.-S. Hua, "Optimizing video search reranking via minimum incremental information loss," in *Proc. ACM Int. Workshop Multimedia Inf. Retr.*, 2008, pp. 253–259.
- [12] F. Richter, S. Romberg, E. Hörster, and R. Lienhart, "Multimodal ranking for image search on community databases," in *Proc. ACM SIGMM Int. Workshop Multimedia Inf. Retr.*, 2010, pp. 63–72.
- [13] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and trecvid," in *Proc. ACM SIGMM Int. Workshop Multimedia Inf. Retr.*, 2006, pp. 321–330.
- [14] C. G. M. Snoek, K. E. A. van de Sande, O. de Rooij, B. Huurnink, J. C. van Gemert, J. R. R. Uijlings, J. He, X. Li, I. Everts, V. Nedovic, M. van Liempt, R. van Balen, F. Yan, M. A. Tahir, K. Mikolajczyk, J. Kittler, M. de Rijke, J. M. Geusebroek, T. Gevers, M. Worring, A. W. M. Smeulders, and D. C. Koelma, "The mediamill trecvid 2008 semantic video search engine," in *Proc. NIST TRECVID Workshop*, 2008, pp. 1–14.
- [15] H.-K. Tan and C.-W. Ngo, "Fusing heterogeneous modalities for video and image re-ranking," in *Proc. ACM Int. Conf. Multimedia Retr.*, 2011, p. 15.
- [15] M. Wang, L. Yang, and X.-S. Hua, "MSRA-MM: Bridging research and industrial societies for multimedia information retrieval," Microsoft, Washington, DC, Tech. Rep. MSR-TR-2009-30, 2009.
- [16] X.-Y. Wei, Y.-G. Jiang, and C.-W. Ngo, "Concept-driven multi-modality fusion for video search," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 1, pp. 62–73, Jan. 2011.
- [17] P. Wilkins, T. Adamek, G. J. F. Jones, N. E. O'Connor, and A. F. Smeaton "Trecvid 2007 experiments at dublin city university," in *Proc. NIST TRECVID Workshop*, 2007, pp. 1–8.
- [18] P. Wilkins, P. Ferguson, and A. F. Smeaton, "Using score distributions for query-time fusion in multimediaretrieval," in *Proc. ACM SIGMM Int. Workshop Multimedia Inf. Retr.*, 2006, pp. 51–60.
- [19] P. Wilkins and A. F. Smeaton, "Weighted data fusion for content-based multimedia information retrieval," in *Proc. Internet Multimedia Search Min.*, Jan. 2010, pp. 1–45.
- [20] R. Yan, A. Hauptmann, and R. Jin, "Multimedia search with pseudorelevance feedback," in *Proc. ACM Int. Conf. Image Video Retr.*, 2003, pp. 238–247.
- [21] R. Yan, J. Yang, and A. Hauptmann, "Learning query-class dependent weights in automatic video retrieval," in *Proc. ACM Int. Conf. Multimedia*, 2004, pp. 548–555.
- [22] T. Yao, T. Mei, and C.-W. Ngo, "Co-reranking by mutual reinforcement for image search," in *Proc. ACM Int. Conf. Image Video Retr.*, 2010, pp. 34–41.
- [23] T. Yao, C.-W. Ngo and T. Mei, "Circular reranking for visual search," in *Proc. IEEE Trans. Image Processing*, vol. 22, no. 4, 2013.
- [24] D. Zhou, S. A. Orshanskiy, H. Zha, and C. L. Giles, "Co-ranking authors and documents in a heterogeneous network," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2007, pp. 739–744.