Visual Reranking: From Objectives to Strategies

Xinmei Tian
University of Science and Technology of China, Hefei, China

Dacheng Tao
University of Technology, Sydney, Australia

A study of the development of visual reranking methods can facilitate an understanding of the field, offer a clearer view of what has been achieved, and help overcome emerging obstacles in this area.

With the rapid development of recording and storage devices, as well as the significant improvement of transmission and compression techniques, the amount of multimedia data (for example, image, video, and audio) on the Web is increasing and the video- and image-sharing sites are becoming more and more popular. There are hundreds of millions of videos on YouTube, Tudou, and Youku. Flickr hosts more than five billion images and Facebook users have uploaded more than 50 billion photos. Many people search and view images and videos on the Web every day. YouTube, for example, serves more than one billion videos every day. As a consequence, efficient and effective multimedia search tools are essential for Web surfing.

There are usually two ways to perform multimedia search: content-based and text-based. In content-based retrieval, which has been used extensively in the past two decades, a user provides example images or video clips and then similar images and videos are returned by querying a visual-representation index in a large-scale database. The content-based method suffers from three disadvantages. The first disadvantage is the well-known semantic gap between low-level visual features and high-level semantic concepts, leading to irrelevant images returned. The second disadvantage is the redundancy of the search result. Content-based image and video retrieval (CBIVR) usually ranks images according to their visual similarities with regard to the query examples. As a consequence, many images that vary slightly from the examples are returned as the top results. Although these near-duplicate images are relevant, they provide insufficient information for users. The third disadvantage is the neglect of user experiences. Example images and video clips essentially required by CBIVR might be unavailable for most users.

Furthermore, with one or few examples, a user's search intention cannot be clearly expressed. If the user provides an example image with a horse on grassland, what does the user indeed want? It could be horse images with various backgrounds, any animal (for example, a goat, a cow, or a lion) on grassland, or something else altogether. Additionally, it could be difficult or even impossible for users to find proper examples to express complex intentions. Searching with textual queries is more natural for users, and it leads to another important search style, that is, text-based multimedia search.

Text-based multimedia search completely relies on indexing the associated textual information of images, such as image tags, webpage filenames, and surrounding text. With textual information, well-understood search techniques can be applied directly to image and video search. Text-based multimedia search is efficient and has been widely used in practical applications. Most image and video search engines, such as Google and Bing, are built around this method. Although text-based multimedia search has the aforementioned advantages, it also suffers from several drawbacks. The first is the mismatching between the images and videos and their associated textual descriptions. The second disadvantage is that the textual representation could be ambiguous because of the influence of polysemy and synonymy. Third, textual information is insufficient to distinguish images of different relevance, which means that some slightly relevant samples will be returned as the results.

To address the problems existing in current multimedia search, visual reranking has become a popular method in recent years. Visual reranking is an integrated framework (see Figure 1) that aims to obtain effective retrieval results efficiently. It leverages the advantages of content-based and text-based retrieval.

1070-986X/11/$26.00 © 2011 IEEE

Published by the IEEE Computer Society

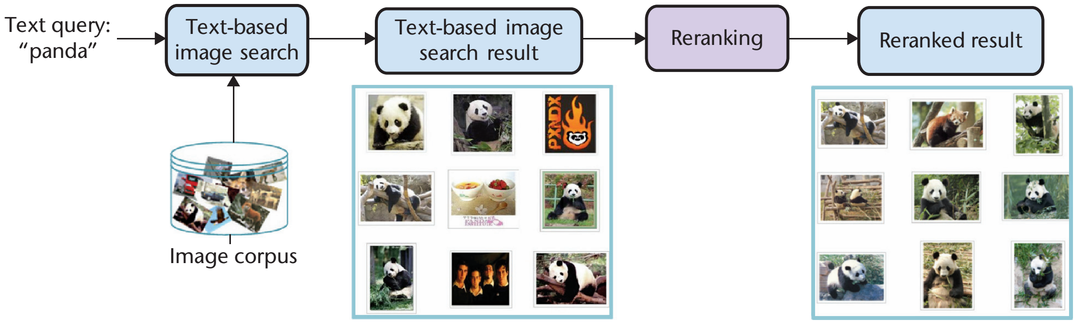

Figure 1. An illustration of a visual reranking framework. First, the text-based search returns the images and video shots releated to the query "panda" from text cue efficiently. Then a visual reranking process is applied to refine the search results by mining visual information. By doing so, more satifactory results are obtained.

In particular, text-based retrieval is first performed to return a rough result efficiently and then efforts are conducted to reorder the samples in the small set to rank the most relevant samples as the top ones for obtaining better search performance. In addition to reordering the image search result generated from text-based search, visual reranking can reorder the initial results derived from the search with other modalities, such as visual, audio, and location information. This article offers a comprehensive survey of recent literature in this area.

Reranking objectives

In multimedia search, the common objective of reranking is to obtain satisfactory search results for providing good search experiences for users. A good search result should return the samples that are relevant and cover a wide range of topics at the same time. That is, the results should be relevant and diverse. According to this point, existing reranking methods can be classified into two categories, relevance-based reranking and diversified reranking (see Figure 2). In relevance-based reranking, higher rank is assigned to more relevant samples without considering diversity. In diversified reranking, the goal is to maximize the result's coverage.

Relevance-based reranking

In the early years of visual reranking, most work focused on improving the quality of

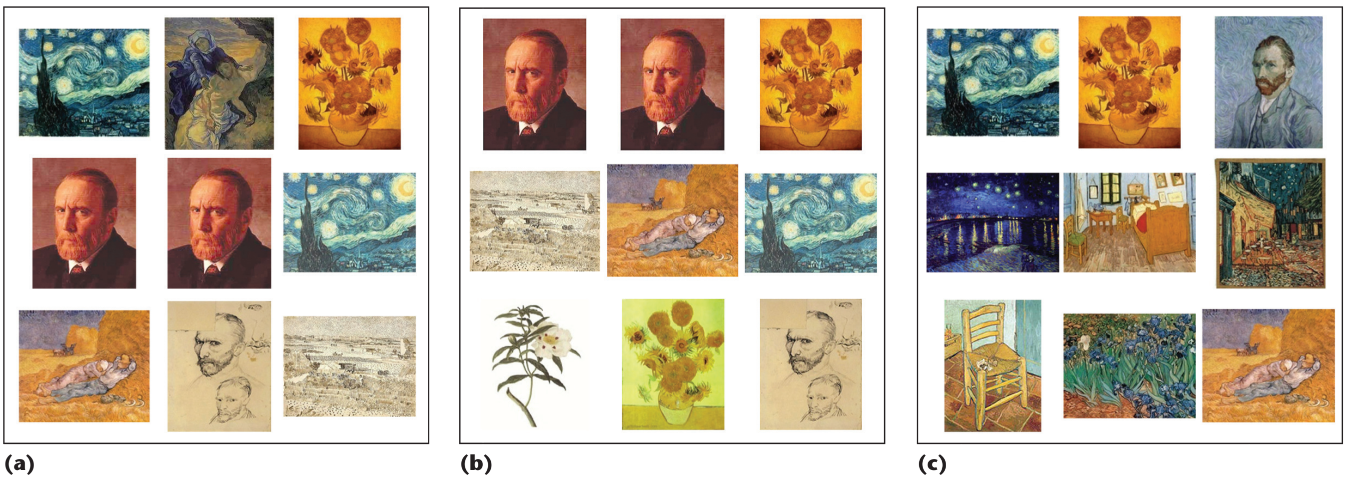

Figure 2. An illustration of the difference between relevance-based reranking and diversified reranking using the query "van Gogh." In (a), (b), and (c), the top-nine images generated by text-based search, relevance-based reranking, and ideal diversified reranking are given. Both the text-based search result and relevance-based reranking result generate near-duplicate results: the two "starry night" paintings in (a) and the two "sunflower" paintings in (b). A disired perfect diversified reranking result, shown in (c), returns the pantings of van Gogh without duplication.

search results from the relevance aspect, that is, boosting the rank of relevant samples.[1-4] There are two commonly used assumptions in relevance-based reranking, visual consistency and ranking consistency. The visual-consistency assumption denotes that visually similar samples are likely to have similar relevance scores and therefore should be adjacently ranked. The ranking-consistency assumption refers to text-based search results, though noisy, reflecting the information obtained from textual cues, and thus should be approximately preserved. Furthermore, top-ranked results returned by initial text-based search are usually more relevant than bottom-ranked results. By leveraging this property, several methods can perform visual reranking via pseudo-relevance feedback.

For relevance measurement, many criteria have been proposed, such as precision, recall, noninterpolated average precision (AP), and normalized discounted cumulated gain (NDCG). The most popular criteria in relevance-based reranking are AP and NDCG. AP averages the precision values obtained when each relevant sample occurs. The AP of top-k ranked sample AP@k is defined in Equation 1,

$$\mathrm{AP@}k = \frac{1}{Z_k}\sum_{i=1}^{k}[p(i) \times rel(i)] \quad (1)$$

where $p(i)$ is the precision at rank i and $rel(i)$ is the binary function on the relevance of the i-th ranked sample with "1" for relevant and "0" for irrelevant. The Z_k is a normalization constant that is chosen to guarantee AP@$k = 1$ for a perfect ranking result. Specifically, $Z_k = k$ if $k < n$, else $Z_k = n$, where n is the total number of relevance samples. Here, the perfect ranking result is derived by ordering images according to their ground-truth relevance labels. In other words, all relevant images are ranked before irrelevant images.

AP is a good measure for the samples labeled with two relevant levels (relevant and irrelevant). However, when the samples have more than two relevant levels (for example, some datasets are labeled with three relevant levels: highly relevant, relevant, and irrelevant), AP cannot be adopted anymore. In this situation, NDCG is widely used in information retrieval to measure the performance when there are more than two relevance levels. For a given query, the NDCG score at rank k in the ranking list l is calculated as

$$\mathrm{NDCG@}k = \frac{1}{Z_k}\sum_{i=1}^{k}(2^{t_i} - 1)/\log(1+i)$$

where t_i is the relevant degree of the i-th image in l and Z_k is a normalization constant, which is chosen to guarantee that the perfect ranking's NDCG@$k = 1$.

The objective of relevance-based reranking is to maximize the relevance of the returned images through reordering. However, since maximizing the relevance of each item in the list is the only objective and each image's relevance is considered independently, the resulting ranking result tends to return a large number of redundant images that convey repetitive information. For example, duplicate, near duplicate, and visually similar images tend to be placed at the top of the list. A ranking list with duplicate or near-duplicate images, although achieving high relevance, provides little information to users.

Diversified reranking

Because most of relevance-based reranking methods rely on visual consistency to perform reranking, visually similar images are often ranked nearby. Near-duplicate images present less information to users, especially in response to queries that are ambiguous, such as "apple." Many researchers have found that users are not very clear on what they want when performing such searches. Thus, a diverse result covering rich topics may meet the various needs of users more effectively and could help them reach their search targets more quickly.

Because search results with rich topic coverage are preferred by users, another genre of reranking, that is, diversified reranking, has been proposed. This method generates a diverse search result by postprocessing the initial text-based retrieval result. Figure 2 shows the top nine images generated by text-based search, relevance-based reranking, and ideal-diversified reranking. In Figure 2b, near-duplicate images (two "sunflower" paintings) are generated by relevance-based reranking methods, while Figure 2c shows a diversified reranking result.

Recently, more emphases has been put on this new kind of reranking method. Diversified reranking maximizes both the number of relevant images and the number of relevant image clusters represented within the top-k results. In one study,[5] the diversity of image search results is achieved by building a retrieval model that relies on the image's associated textual features, such as the tag, title, and description. In another study,[6] the images are first clustered via various clustering algorithms and then the desired diverse result is formed by picking up one representative image from each cluster.

Evaluation measures used in relevance-based reranking cannot be used in diversified reranking, so specific evaluation measures must be developed to measure the performance of diversified reranking. These measures should take both relevance and diversity into consideration. One commonly used measure in diversified reranking is based on cluster recall and precision at a fixed rank of k, that is, $CR@k$ and $p@k$. The cluster recall measured at rank k is defined as

$$CR@k = \frac{clusters(k)}{tc}$$

where $clusters(k)$ is the number of clusters covered by the top-k ranked images and tc is the total number of clusters for the particular query. The CR corresponds to the S-recall proposed in text retrieval.[7] The final measure is defined as the F-score, that is, the harmonic mean of precision p and CR,

$$f = \frac{2P \cdot CR}{P + CR}$$

This measure uses CR to measure diversity and precision to measure relevance. However, the images in search results are often in hierarchical structure and CR treats each cluster equally. Generally, covering a more popular or important topic or cluster is preferable to covering a rare topic. Therefore, it's reasonable to take topic or cluster importance into consideration when measuring performance. User studies are the most direct way to measure satisfaction with the diversified reranking results. However, it's time-consuming and requires too much human effort.

Reranking features

According to the features used, visual reranking methods can be categorized into three groups: textual, high-level visual, and low-level visual.

Textual features

Although most reranking methods reorder the initial text search result by introducing visual information, there are several research projects that reorder the initial text search result by further exploiting the textural feature. W.-H. Lin et al.[8] used the HTML documents associated with the images returned from the text-based image search engine to build relevance models. By exploiting the global information of the documents, the images are reordered according to the probability estimated by the refined relevance model. The diversified reranking method proposed in another project[5] also used the textual information associated with the images, such as tags, titles, and descriptions, to build language models to generate diverse results.

The textual-feature reranking methods translate visual reranking to conventional textual search reranking and then leverage the techniques already developed in that domain. The drawback of this approach is that essential visual features are neglected. As a consequence, the problems encountered in initial text-based search still exist. To overcome this drawback, many reranking methods rely on other modality features, such as visual features, to boost search results by introducing complementary information.

High-level visual features

Of the high-level visual features methods, such as concept detection, annotation, facial detection, and anchor detection, the most frequently used is concept detection. There are usually three main steps in concept-detection reranking. First, by analyzing textual queries, several query-related concepts are selected from a predefined concept set. Then, by using trained concept detectors, the concept detection scores of the query-related concepts for each image are obtained. Finally, reranking is conducted by using the obtained concept-detection scores via several methods, such as ListNet.[9] For the predefined concept set in the first step, the Large Scale Concept Ontology for Multimedia provides a set of concept detectors for 449 concepts. Most of the concept-detection reranking methods rely on this concept set.[9]

For reranking in concept space, the key problem is to determine the query-related concepts, that is, the query expansion. Natsev et al.[10]

reviewed query expansion for visual reranking, categorizing the query-expansion methods into two classes: text based and visual concept based. In each class, the methods are further divided into language-specific lexical approaches, corpus-specific statistical approaches, and query-specific statistical approaches. Details can be found elsewhere.[10] High-level visual-feature reranking methods achieve good performance for specific queries, such as facial detection for person-related queries. On the other hand, these methods are limited to these small-amount queries.

Low-level visual features

Although high-level visual features can boost the reranking performance to some extent, both the scalability and accuracy are still far from satisfactory for large-scale practical use. To avoid accumulating the errors introduced in concept detection, most reranking methods are directly conducted in the low-level visual feature space. The most frequently used low-level visual feature includes local features (for example, scale-invariant transform), and global features (for example, color moment, Wavelet texture, and edge-direction histogram).

Global features characterize the content of images by regarding each image as a whole. Commonly used global features in reranking include the color (for example, color moments), the edge (for example, edge-direction histogram), and the texture (for example, wavelet texture). Different features describe images from different aspects and they are complementary with each other. Therefore, they are often fused to achieve better performance. There are usually two ways to fuse multimodal global features. One is early fusion in which the multimodal features are concatenated into one feature vector. Then this long feature vector is used for reranking. For example, Tian et al. represented images with 428D global visual features consisting of 225D color moment, 128D wavelet texture, and 75D edge distribution.[11] Leuken et al. alternatively proposed to use dynamic feature-weighting to measure the visual similarity from six visual features, including color, shape, and texture.[6] In this method, the images are clustered according to the obtained multimodality similarities. The other method is later fusion in which reranking is first performed by considering each single feature independently. Afterwards, all reranking results derived with single features are fused to get the final reranking.[10,12]

Local features describe images by considering them as a collection of local patches. A popular local feature is the bag-of-visual-words representation. A set of local descriptors are first extracted from the images. Then a codebook is generated by clustering all the local descriptors. Each entry in the codebook, each so-called visual word, corresponds to the center of each cluster. Finally, by quantizing the local descriptors in image I_i into visual words, I_i can be represented as a vector based on term frequency. Jing and Baluja used scale-invariant transforms to measure the visual similarity to construct a graph for reranking.[2] Yao et al. represented images as a histogram consistent with the bag-of-visual-words method, and then calculated the visual similarity between images according to the cosine distance between their corresponding histograms.[3] Hsiao and Chen adopted the bag-of-visual-words method for image representation and then used a language model to rank images.[13]

Global features and local features characterize the visual content of the images from different aspects and complement each other. More effective features can be derived by combining the local and global features to possess the advantages of both. Textual, high-level visual, and low-level visual features show their strength in different ways. Richter et al. proposed to model the similarity between images from multimodal cues,[14] specifically measuring similarity on the basis of low-level visual features and textual features (user tags). The good performance reported in this research should encourage work on more reranking approaches that can seamlessly fuse multimodal features. In addition, methods that can automatically determine the most suitable feature for specific queries would be very useful. State-of-the-art dimension reduction and distance-metric-learning techniques[15] could also be introduced to exploit effective visual features for visual-content representation.

Reranking strategy

At the early stage of improving text-based search results by incorporating visual information, simple linear combination was used to combine the search results derived from textual and visual cues.[10,12] Later, more effective tools were developed to fuse the textual and visual cues more elaborately with sophisticated tools, such as support vector machine,[12] ListNet,[9] information bottleneck principle, random walk,[1] PageRank,[2] and so on. These reranking methods can be divided into three categories: clustering, classification, and graph, as illustrated in Figure 3.

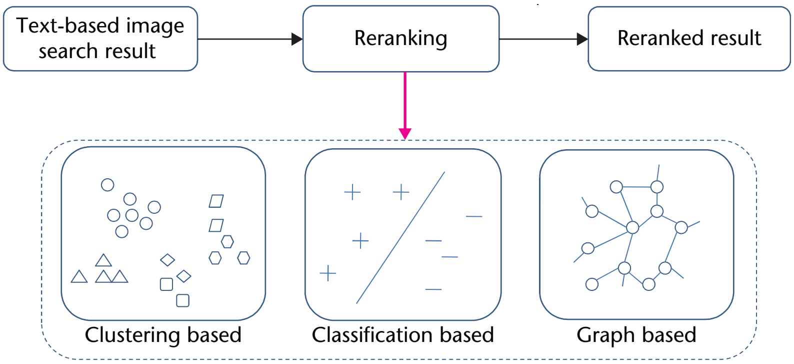

Figure 3. An illustration of three different reranking strategies: clustering, classification, and graph reranking.

Clustering

Clustering methods are based on the observation that query-relevant images often share high visual similarity. By using various clustering algorithms, this kind of method reorganizes the initial text search result by grouping visually similar samples together. In Hsu et al.,[1] each sample is given a soft pseudo label according to the initial text search result. Then the information bottleneck principle is adopted to find optimal clustering. Information bottleneck is motivated from Shannon's rate-distortion theory and it seeks a clustering result that minimizes the loss of mutual information between features and label. Hsu et al. manually set the number of clusters to ensure each cluster contains about 25 items.[1] Reranking results are achieved by ordering the clusters according to the cluster-conditional probability first and then ordering the samples within a cluster on the basis of their local feature density approximated via kernel density estimation. This method improved the search performance from 0.169 to 0.204, 0.087 to 0.102, and 0.087 to 0.107 in Trecvid 2003–2005, respectively. It achieves good performance on the named-person queries.[1] But it is limited to those queries that have significant duplicate characteristic.

Jing et al.[16] reorganized the Web image search results into clusters. They first identify query-related semantic clusters by using Web page search, and then use the derived query-related cluster names as queries for an image search engine. A user study showed that these results were preferred to those generated by Google. This method leverages the power of text search, but it also means that Web image search results for subclusters still rely on text-based search. Finally, Leuken et al. grouped the image search results according to the dynamic visual similarity derived from three heuristic-clustering algorithms: folding, max-min, and reciprocal election.[6]

Clustering methods are suitable for queries that have obvious near-duplicate images in the initial text-based results. For these queries that return visually diverse images without salient patterns, this kind of method cannot achieve good performance. Besides, automatically determining the number of clusters is still an open problem.

Classification

In the classification method, reranking is simplified as a binary classification problem. There are normally three steps: select the pseudo-positive and pseudo-negative samples from the initial text-based search results; train a classifier using the selected samples; and reorder the samples according to the relevance scores predicted by the trained classifier.

For the first step, pseudo relevance feedback (PRF) is typically used to select training samples. PRF is a concept introduced from text retrieval. It assumes that a fraction of the top-ranked documents in the initial search results are pseudo-positive. Alternatively, Yan et al. used the query images or example video clips as the pseudo-positive samples.[12] The pseudo-negative samples are selected from either the least-relevant samples in the initial ranking list or the database, with the assumption that few samples in the database are relevant to the query.

In the second step, various classifiers, such as support vector machine[12] and ListNet,[9] can be used. Although these classifiers are effective, sufficient training data is required to achieve satisfactory performance because a lot of parameters must be estimated. However, in visual reranking, the training data usually obtained via PRF is noisy due to the imperfect text-based search result and insufficient, restricting the performance of this kind of method for real-world applications.

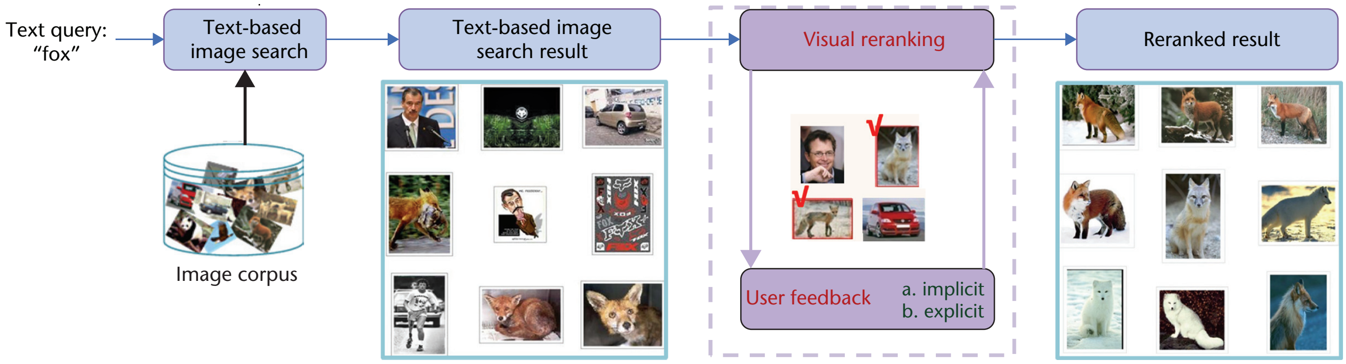

Figure 4. An illustration of reranking with user feedback. After the text-based search result is obtained, user feeback is introduced to help perform reranking more effectively. The feedback information can either be explicitly obtained from user interaction or implicitly collected from log files.

Graph

In the graph methods,[1-4,14] a graph is constructed to mine the relations between the images. The graph is constructed with the samples as the nodes and the edges between them being weighted by visual similarity. Then, reranking is performed on the graph by propagating the ranking scores through the edges. In graph methods, the relationships of all samples are represented by the graph. Therefore, the graph construction plays the key role in this kind of method.

Hsu et al. formulated the reranking as a random walk over the graph.[1] To leverage the text search result, a dongle node is attached to each sample with the value fixed to be the initial text-ranking score. The stationary probability of the propagation process is adopted as there ranked score. Hsu et al. used this method to improve the Trecvid 2005 text search result and reported overall story-level improvement from 0.204 to 0.271. Jing and Baluja applied the well-known PageRank algorithm to rerank Google image search by treating images as documents and their visual similarities as probabilistic hyperlinks.[2] This method achieves good performance on product and landmark queries.

Yao et al. proposed co-reranking by conducting a random walk on two graphs for visual and textual information.[3] In addition, Cao et al.[7] extended VisualRank[2] to rank and cluster the images. The experiments conducted on the Cross Language Evaluation Forum 2008 dataset show that it outperforms VisualRank both in precision and topic recall. Tian et al. proposed a general framework for graph reranking that formulated reranking to maximize the visual consistency over the graph and minimize the ranking distance between the reranked and initial text-based search result.[4] The random walk methods proposed in other research[1,2] can be unified into this framework, and can achieve the best performance of graph-based reranking methods on Trecvid 2006–2007.

User search intention

Due to the ambiguity of query words, the images returned by text-based search might belong to several topics. For example, with the query "apple," the returned images include the fruit and products of the company. Without user interaction, there is no way to decide which topic is preferred by the user. The diversified reranking methods solve this problem by providing diverse results to help users access their targets quickly. Diversified reranking is a good solution for solving this problem if user interaction is unavailable. When user interaction is available, reranking system can resort to user feedback to learn search intentions. With the user's search intention known, reranking can generate personalized reranking results in which only interesting images remain and uninteresting images are removed.

Recently, several projects have been undertaken in this area. Cui et al. proposed a method named IntentSearch, which interacted with the users by allowing them to specify one example image from the initial text-based search result.[17] Then the remaining images are ranked according to their relevance. Due to the semantic gap, one image is usually insufficient to express complex search intentions. Tian et al. proposed a general active reranking framework that iteratively obtains feedback to learn user's search intention.[11] IntentSearch can be regarded as a simplified active reranking method with only one relevant image labeled by the user. Hsiao and Chen proposed an intention-focused active reranking method to deal with retrieval.[13] In this method, they added an intention region to confirm feedback and avoid background noise. A general framework of reranking with relevance feedback is shown in Figure 4.

In addition to using explicit feedback, implicit feedback can be used for visual reranking. Although implicit feedback is noisy, it usually can be obtained conveniently and does provide useful information. Clickthrough, a typical implicit feedback collected from log files, has proven to be useful in improving the performance of search engines for Web documents searching. Recently, Jain and Varma introduced clickthrough data into image search reranking to build query-dependent image reranking models.[18] It performed well and improved Bing image search (from an NDCG of 0.685 to 0.769). The success of clickthrough for visual reranking encourages us to collect and evaluate variant implicit user feedback in visual reranking.

Conclusions

Although a lot of work has been conducted on visual reranking, there are still many problems. From the objective perspective, most of the diversified reranking methods are conducted in two steps: reranking the samples without considering diversity and then heuristically generating a visual list as a post-processing step. We believe there should be more research on developing reranking algorithms that can directly generate diverse results by optimizing both relevance and diversity jointly. In addition, researchers working in this area need a good evaluation measure that takes the hierarchical structures of the samples into consideration.

Visual features can play a more critical role in visual reranking. On one hand, novel methods can be developed to effectively combine variant visual features that complement each other. On the other hand, with comprehensive visual features, state-of-the-art dimension reduction and metric-learning techniques could be used to exploit visual features more effectively and further boost reranking performance. Especially when the user feedback is available, distance metric learning and dimension reduction could become even more effective at encoding user feedback information in the feature space.

Finally, because implicit feedback is useful and has the advantage that it can be collected at low cost and in large quantities, new reranking algorithms should exploit implicit feedback information to deal with the noise. **MM**

References

1. W.H. Hsu, L.S. Kennedy, and S.-F. Chang, ''Reranking Methods for Visual Search,'' *IEEE Multimedia,* vol. 14, no. 3, 2007, pp. 14-22.
2. Y. Jing and S. Baluja, ''Visualrank: Applying PageRank to Large-Scale Image Search,'' *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, 2008, pp. 1877-1890.
3. T. Yao, T. Mei, and C.-W. Ngo, ''Co-Reranking by Mutual Reinforcement for Image Search,'' *Proc. ACM Int'l Conf. Image and Video Retrieval* (CIVR), ACM Press, 2010, pp. 34-41.
4. X. Tian et al., ''Bayesian Video Search Reranking,'' *Proc. ACM Int'l Conf. Multimedia,* ACM Press, 2008, pp. 131-140.
5. R. Zwol et al., ''Diversifying Image Search with User Generated Content,'' *Proc. ACM Int'l Conf. Multimedia Information Retrieval,* ACM Press, 2008, pp. 67-74.
6. R.H. Leuken et al., ''Visual Diversification of Image Search Results,'' *Proc. Int'l Conf. World Wide Web,* ACM Press, 2009, pp. 341-350.
7. L. Cao et al., ''Rankcompete: Simultaneous Ranking and Clustering of Web Photos,'' *Proc. Int'l Conf. World Wide Web,* ACM Press, 2010, pp. 1071-1072.
8. W.-H. Lin, R. Jin, and A.G. Hauptmann, ''Web Image Retrieval Re-Ranking with Relevance Model,'' *Proc. IEEE Int'l Conf. Web Intelligence,* IEEE CS Press, 2003, pp. 242-248.
9. Y.-H. Yang et al., ''ContextSeer: Context Search and Recommendation at Query Time for Shared Consumer Photos,'' *Proc. ACM Int'l Conf. Multimedia,* ACM Press, 2008, pp. 199-208.
10. A. Natsev et al., ''Semantic Concept-Based Query Expansion and Re-Ranking for Multimedia Retrieval,'' *Proc. ACM Int'l Conf. Multimedia,* ACM Press, 2007, pp. 991-1000.
11. X. Tian et al., ''Active Reranking for Web Image Search,'' *IEEE Trans. Image Processing,* vol. 19, no. 2, 2010, pp. 805-820.
12. R. Yan, A. G. Hauptmann, and R. Jin, ''Multimedia Search with Pseudo-Relevance Feedback,'' *Proc. ACM Int'l Conf. Content-Based Image and Video Retrieval,* ACM Press, 2003, pp. 238-247.
13. J.-H. Hsiao and M.-S. Chen, ''Intention-Focused Active Reranking for Image Object Retrieval,'' *Proc. ACM Int'l Conf. Information and Knowledge Management,* ACM Press, 2009, pp. 157-166.
14. F. Richter et al., ''Multimodal Ranking for Image Search on Community Databases,'' *Proc. ACM Int'l Conf. Multimedia Information Retrieval,* ACM Press, 2010, pp. 63-72.

15. Y. Yang et al., "Harmonizing Hierarchical Manifolds for Multimedia Document Semantics Understanding and Cross-Media Retrieval," *IEEE Trans. Multimedia,* vol. 10, no. 3, 2008, pp. 437-446.
16. F. Jing et al., "IGroup: Web Image Search Results Clustering," *Proc. ACM Int'l Conf. Multimedia,* ACM Press, 2006, pp. 377-384.
17. J. Cui, F. Wen, and X. Tang, "Real Time Google and Live Image Search Re-Ranking," *Proc. ACM Int'l Conf. Multimedia,* ACM Press, 2008, pp. 729-732.
18. V. Jain and M. Varma, "Learning to Re-Rank: Query-Dependent Image Re-Ranking Using Click Data," *Proc. Int'l Conf. World Wide Web,* ACM Press, 2011, pp. 277-286.

Xinmei Tian is with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. Her research interests include multimedia analysis and retrieval, computer vision and machine learning. Tian has a PhD in electronic engineering and information science from the University of Science and Technology of China. Contact her at xinmeitian@gmail.com.

Dacheng Tao is a professor of computer science with the Centre for Quantum Computation and Information Systems and the Faculty of Engineering and Information Technology at the University of Technology, Sydney. His research interests include applying statistics and mathematics to data analysis problems in data mining, computer vision, machine learning, multimedia, and video surveillance. Tao has a PhD in computer science and information systems from the University of London. He received the best theory/algorithm paper runner up award at the 2007 IEEE International Conference on Data Mining. Contact him at dacheng.tao@uts.edu.au.