AN EFFICIENT GRAPH-BASED VISUAL RERANKING

Chong Huang[†], *Yuan Dong*[†], *Hongliang Bai*[‡], *Lezi Wang*[†], *Nan Zhao*[†], *Shusheng Cen*[†],*Jian Zhao*[§]

[†]Beijing University of Posts and Telecommunications,100876, P.R.China
[‡]France Telecom Research & Development - Beijing, 100190, P.R.China
[§]Patent Examination Cooperation Centers of the Patent Office, SIPO, Beijing, P.R.China
{huangchong661100, yuandong, wanglezi, nanzhao, censhusheng}@bupt.edu.cn
hongliang.bai@orange.com, zhaojian_2@sipo.gov.cn

ABSTRACT

The state of the art in query expansion is mainly based on the spatial information. These methods achieve high performance, however, suffer from huge computation and memory.

The objective of this paper is to perform visual reranking in near-real time regardless of the spatial information. We explore a graph-based method proposed by [1] as our confident sample detection baseline, which has been proved successful in achieving high precision. In addition, a novel maximum-kernel-based metric function is introduced to rerank the images in the initial result.

We evaluated the method on the standard *Paris* dataset and a new *Francelandmark* dataset. Our experiments demonstrate that the algorithm has great value on practicality because of its good performance, easy implementation, and high computational efficiency.

Index Terms— Query Expansion, Reciprocal Neighbor, Confident Sample, Maximum-kernel-based

1. INTRODUCTION

With the development of Internet and computer technology, the content-based visual retrieval has been researched over several years. Many works have constructed the standard framework [2, 3, 4]: each image is represented using bag-of-word (BoW), and images are sorted using term frequency-inverse document frequency (tf-idf) computed efficiently via an inverted index.

Unfortunately, the traditional content-based search fails to perform perfectly. There are several factors accounting for it: inappropriate metrics for descriptor comparison; noisy descriptors; feature detection drop-out; or loss due to descriptor quantization [5]. To address these problems, query expansion is introduced to the visual reranking.

Query expansion, originated from the text retrieval literature, has received attention in the computer vision over the past years [6]. Given an initial ranking list and its corresponding set of visual feature $R = [r_1, r_2, \cdots, r_M]$ and $f = [f_1, f_2, \cdots, f_M]$ respectively, where M is the number

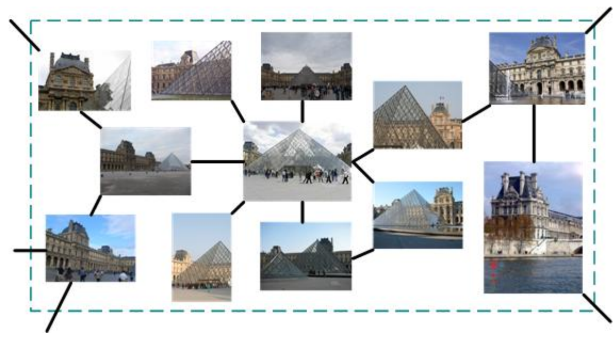

Fig. 1. Image graph. The images related with the same object are connected to a subgraph.

of the candidates, the high-ranking images from the original query are reused as the requery. which can be described as follows:

$$q' = g(q, R, f)$$

where q is the original query and q' is a set of requery.

$$R' = h(q', q, f)$$

where R' is the reranked result.

Depending too much on the initial list is the principal limitations of the query expansion, because a form of blind relevance feedback may fail if false-positive images are included in q'. Many works [4, 7, 8, 9] have made contribution to the query expansion. [6] proposes that strong spatial constraints between the query image and each result allow accurately verifying each return, such as RANSAC.

However, these methods often suffer from huge computation and memory. It is not practical in the near-real time search. Feature augmentation is another natural complement to the query expansion. The database-side feature augmentation is proposed that images in the database are augmented offline with spatially verified visual words [5]. On the other hand, [10] introduces k-nearest neighbors(k-NN) of query for automatically refining the initial list, which meet state-of-the-art retrieval performance on the several public databases.

Inspired by the spirit of image augmentation and k-NN, our approach explores graph-based method to rerank the initial list. It includes two parts: offline indexing reciprocal neighbors, and online reranking the initial list.

The rest of this paper is organized as follows. Section 2 describes our system including offline indexing and online reranking. Experiments are conducted to demonstrate the effectiveness and efficiency of the algorithm in Section 3. Section 4 concludes and proposes the future work.

2. GRAPH-BASED RERANK

2.1. Offline Indexing

The framework of offline indexing is shown in Fig. 2.

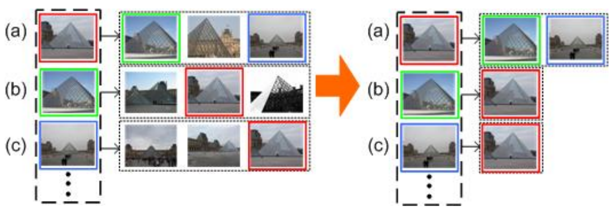

Fig. 2. Framework of offline indexing. Image *a* and image *b* belong to their respective *k* nearest neighbors(indicated in the red box and the green box), so they meet the condition of reciprocal neighbors. Only the images which are reciprocal neighbors images are kept.

For the dataset, assumed that the images containing the same view of the object are connected, each image in the dataset would seek the latent connected candidates based on the reciprocal neighbors relation. As discused in [11, 12], the reciprocal neighbor relation is defined as:

$$R_k(i,i') = i \in N_k(i') \wedge i' \in N_k(i) \tag{1}$$

where $N_k(i)$ is the set of the k nearest neighbors of image i. The reciprocal neighbor is a reliable guarantee to evaluate the visual similarity between two images. We use the reciprocal-neighbor graph $G = (V, E, W)$, where V is the set of the images, E is the set of edges connecting images, and W is defined as

$$w(i,i') = \begin{cases} \frac{N_k(i)\cap N_k(i')}{k} & if(i,i') \in R_k(i,i') \\ 0 & otherwise \end{cases} \tag{2}$$

2.2. Online Reranking

The online reranking consists of two steps: The first step is to find the confident images which contribute to the query expansion, and the second step is to rerank the images by using the detected samples. The framework is shown in Fig. 3.

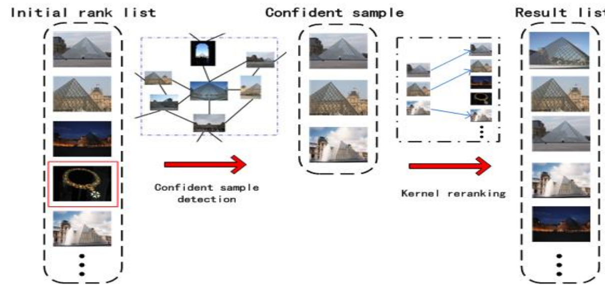

Fig. 3. Framework of online searching. After the confident sample detection and kernel reranking, the irrelevant image(indicated in the red box) is filtered out.

2.2.1. *Confident Sample Detection*

The confident sample detection is based on the graph-based ranking. More concretely, we follow the architecture of the previous work [1] for a baseline retrieval system, which has been proven successful in the high precision. The objective is to search for a subgraph G' from G maintaining the maximum density in Eqn. 3. Then, the initial list is reranked according to the time of insertion into subgraph G'.

$$G' = \underset{G'=(V',E',W)\subset G: q\in V'}{\operatorname{argmax}} \frac{\sum_{(i,i')\in E'} w(i,i')}{|V'|} \tag{3}$$

For certain queries, the principal limitation of the method is that it could not ensure high recall if few reciprocal neighbors can be found and built. Even so, the graph-based method can be used to detect the confident samples, because the confident samples must maintain high precision. The subgraph starts with query and most confident samples are inserted into the subgraph successively. An approximate solution is adopted to solve the Eqn. 3 as Algorithm 1.

2.2.2. *Kernel-Based Reranking*

So far, the confident samples are detected. A simple and novel kernel-based metric function is introduced to rerank the images in the initial list.

Maximum-kernel-based expansion. Because the similarities between the relevant images are generally higher than those of the irrelevant images, the relevant images receive higher supports from the confident samples while the irrelevant images receive lower supports [13]. In addition, the rank in the confident sample reflects the relevance level against the query. Therefore, it could reduce the impact from false positive images by weighting the confident images. Considering the convention that the image with smaller metric would rank higher, we formalize the metric function as follows.

$$s_i = min\left\{ \beta^{r_n} \frac{\|f_i - f_n\|_2^2}{\sigma_n^2} \mid n = 1, 2, .., N_c \right\} \tag{4}$$

Algorithm 1: Confident sample detection

Input:
q, database $D = \{i_1, i_2, \cdots, i_M\}$, parameter k, N_c
Initialization:
confident nodes $I = \{q\}$, edge nodes $E = \{\varnothing\}$,
reciprocal neighbor $R = \{r \in D \mid R_k(q, r)\}$,
outer nodes $O = \{D \setminus R\}$, $N_t = 1$
while $N_t \leq N_c$ **do**
  $E \leftarrow \{E \cup R\}$

$$e = \underset{e \in E}{\operatorname{argmax}} \sum_{i \in I} w(i, e) \alpha^{max(r_i, r_e)}$$

  where α is a constant ranging from 0.5 to 0.8, and $r_i(r_e)$ is the initial ranking of $i(e)$.
  $I \leftarrow \{I \cup \{e\}\}$
  $E \leftarrow \{E \setminus \{e\}\}$
  $R \leftarrow \{r \in O \mid R_k(e, r)\}$
  $O \leftarrow \{O \setminus R\}$
  $N_t + +$
end
Output: I

where r_n and f_n is the order and the feature vector of the nth images in the confident samples, β is set as 0.99, and σ^2 is set as

$$\sigma_n^2 = \sum_{m=1}^{M} \|f_n - f_m\|_2^2$$

2.3. Complexity and Scalability

We analyze the complexity and scalability of our graph-based reranking algorithm. In the offline indexing, the time complexity of reciprocal-neighbor graph is $O(M^2)$, where M is the size of the database. In the online reranking, if N_c confident samples are detected, the complexity ranges from $O(N_c)$ to $O(k \lg N_c)$, and the running time is about 1 second in the experiment. The number of reciprocal-neighbors determines the memory consumption. After the confident samples are found, the kernel-based reranking takes $O(LN_c)$ to re-score the images in the initial list where L is the length of the initial list.

3. EXPERIMENTS

3.1. Dataset

To evaluate our system, experiment are conducted on the two databases - only the *Paris* database, and the *Paris* + *Francelandmark*. *Francelandmark* includes some images crawled from *Flickr*, Bing and Google using queries of famous 78 France landmarks and 24 artworks.

The ***Paris*** [14] includes 6,391 images collected from *Flickr* by searching for particular Paris landmarks. There are 55 images extracted from dataset as the query. The retrieval performance is measured by mAP(mean Average Precision). In addition, the precision at top n ($p@n$) is also selected as the evaluation of users experience, since users often only concern about the first screen of retrieval results.

The ***Francelandmark*** contains 86,717 images in total, which gets closer to the authentic application. The performance is evaluated by the precision at top n candidates. We choose two groups of queries to simulate the real conditions:

- Low Precision (LP): 25 queries where the the precision at top 25 candidates is lower than 30%.
- High Precision (HP): 25 queries where the the precision at top 25 candidates is high than 70%.

Fig. 4. A random samples from the *Francelandmark* dataset

3.2. Methods

As for the visual features, we choose the Harris Laplace detector and SIFT as the descriptor. Then, a random collection of $40M$ descriptors are sampled to learn a $1M$ codebook with Approximate K-means(AKM). Each image is represented by the BoW vector. Several different query expansion methods are compared as follows.

Query Expansion Baseline(QEB). This method assumes the top N candidates to be confident and averages the similarities computed from the entire result image and requery. The N is tuned as 20 to achieve the best performance.

Maximizing Weighted Density(MWD). This method[1] reranks the images according to their time of insertion into subgraph while maximizing weighted density.

Maximum-Kernel-based Expansion(MKE). Confident samples are detected based on the **Algorithm 1** and then the Eqn. 4 is performed.

GroundTruth-based Expansion(GTE). The method picks parts of relevant images from the groundtruth to enable controlled construction of expanding queries. The N is set as 160 in which the MKE performs best.

3.3. Evaluation

Table 1 shows that for the *Paris* dataset, the performance of all methods are better against the initial rank.

Table 1. The mAP (in %) on the *Paris* dataset

INIT	QEB	MWD	MKE	GTE
56.38	63.19	70.52	72.87	82.61

As for QEB, if the top N candidates for each query happened to be correct, the refined list would be better than initial list. However, in fact this can be dangerous because the high-ranking images may include the noise. MWD performs significantly better than blindly choosing high-ranking images for expansion. One of the strengths is its efficiency to got high precision in returning the augmenting images as Fig. 6. This method improves in a limited range since it may be unable to find sufficient reciprocal neighbors for some query. GTE could be regarded as the upper bound of query expansion. When the re-query are sufficient and correct, the performance could reach up to 82.61%.

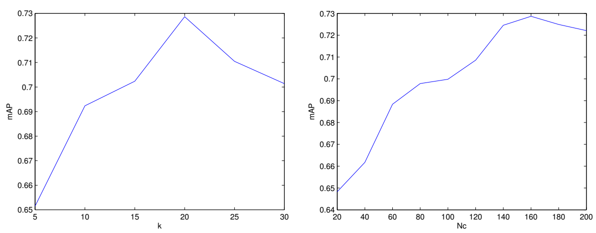

Fig. 5. The mAP curves of MKE in terms of the parameter k (left) and N_c (right) on the *Paris*.

Our method(MKE) improves on the MWD, by adding new constrains and kernel-based metric function. The parameter k and N_c have great influence on the performance. The size of k concerns the strict degree about choosing the reciprocal neighbors. If the reciprocal neighbors are selected too rigorously, the image and its reciprocal neighbors look pretty much the same so as to lose the augmentability. But also, oversize k would render introducing noisy neighbors. In terms of N_c, its role is similar with k, and the difference is that N_c is related to sufficiency and precision of the confident samples. As the Fig. 5 illustrates, the mAP is highest by up to 72.87% when the k and N_c are set as 20 and 160, respectively.

Fig. 7 shows the MKE generously outperforms other methods in the both groups. Even though it is difficult to rank only using visual features in the low precision group, the result could be improved by augmenting the query-related images. In addition, our method is unsupervised and needs not to train the classifier in advance. Therefore, it is suitable for the real-time visual retrieval.

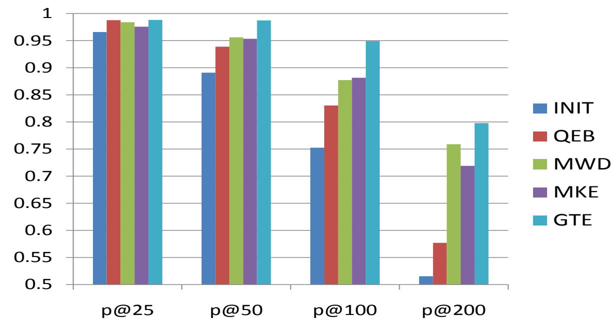

Fig. 6. Comparison of $p@n$(precision at top n) on the *Paris* dataset.

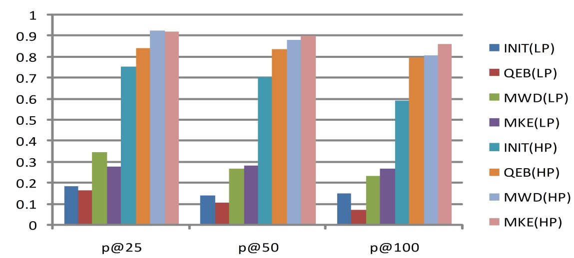

Fig. 7. Comparison of $p@n$(precision at top n) on the *Francelandmark*.

4. CONCLUSION

In this paper, a graph-based query expansion is presented which is suitable for the real-world application. The method absorbs the strength of [1] that achieves high precision, and combines with the kernel-based function to offset its low recall. From the experiment result, we could conclude that this method is efficient in the visual reranking. The second contribution is that we present a new public dataset *Francelandmark* which would facilitate the real-world visual reranking. In the future, we will try to turn spatial verification into graph-based problem so that the graph theory could be introduced to solve it.

Acknowledge The work is sponsored by co-research project SEV01100474 between France Telecom R&D Lab Beijing and Beijing University of Posts and Telecommunications, and Graduate Innovation funding of SICE, BUPT 2011.

5. REFERENCES

[1] Shaoting Zhang, Ming Yang, Timothee Cour, Kai Yu, and Dimitris N. Metaxas, "Query specific fusion for image retrieval," *In Proc. ECCV*, 2012.

[2] D. Nister and H.Steweniius., "Scalable recognition with a vocabulary tree," *In Proc. CVPR*, 2006.

[3] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retreval with large vocabularies and fast spatial matching," *In Proc. CVPR*, 2007.

[4] J. Sivic and A. Zisserman, "A text retrieval approach to obect matching in videos," *In Proc. ICCV*, 2003.

[5] R. Arandjelovi and A. Zisserman, "Three things everyone should know to improve object retrieval," *In Proc. CVPR*, 2012.

[6] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval," *In Proc. ICCV*, 2007.

[7] T.Quack, V. Ferrari, and L. Van Gool, "Video mining with frequent itemset configurations," *In Proc. CIVR*, 2006.

[8] C. Schmid and R. Mohr, "Combining greyvalue invariants with local constraints for object recognition," *In Proc. CVPR*, 1996.

[9] D. Tell and S. Carsson, "Combining appearance and topology for wide baseline matching," *In Proc. ECCV*, 2002.

[10] Xiaohui Shen, Zhe Lin, Jonathan Brandt, Shai Avidan, and Ying Wu, "Object retrieval and localization with spatially-constrained similarity measure and k-nn reranking," *In Proc. CVPR*, 2012.

[11] C. Schmid, H. Harzallah, and J. Verbeek, "Accurate image search using the contextual dissimilarity measure," *IEEE Transactions, Pattern Analysis and Machine Intelligence*, Jan. 2010.

[12] Qin Danfeng, Stephan Gammeter, Lukas Bossard, Till Quack, and Luc VanGool, "Hello neighbor: accurate object retrieval with k-reciprocal nearest neighbors," *In Proc. CVPR*, 2011.

[13] Nobuyuki Morilka and Jingdong Wang., "Robust visual reranking via sparsity and ranking constraints," *ACM International Conference on Multimedia (ACMMM11)*, 2011.

[14] "http://www.robots.ox.ac.uk/ vgg/data/parisbuildings," .