Asian Journal of Convergence in Technology *Volume II Issue III*
Issn No.:2350-1146, I.F-2.71

Enhancing Visual Search by Using Image Re-ranking

Amit A. Yadav
ME Computer Science & Engineering
Annasaheb Dange College of Engg. & Technology,
Ashta, Sangali, Maharashtra
email: amityadavsir@yahoo.com

Prof. Tamboli A. S.
Assistant Professor, Information Technology
Annasaheb Dange College of Engg. & Technology,
Ashta, Sangali, Maharashtra
email: ajs_it@adcet.in

Abstract **- The existing web image search engines, including Bing, Google, and Yahoo, retrieve and rank images mostly based on surrounding text features[8][12]. Image redundancy is still a problem area concerned. It is difficult for them to interpret users search intention only by query keywords and this leads to ambiguous and noisy search results which are far from satisfactory. It is important to use visual information in order to solve the ambiguity in text-based image retrieval. In this paper, we have proposed a novel image search approach. It only requires the user to click on one query image with minimum effort and images from a pool retrieved by text-based search are re-ranked based on both visual and textual content.**

Index Terms – Text based search, Adaptive similarity, Keyword expansion, Visual query expansion, Image Re-ranking.

I. INTRODUCTION

Information retrieval is the activity of obtaining information resources relevant to an information need from a collection of information resources. Searches can be based on metadata. An information retrieval process begins when user enters a query into the system. Queries are formal statements of information needs, for example search strings in web search engines. In information retrieval a query does not uniquely identify a single object in the collection. Instead, several objects may match the query, perhaps with different degrees of relevancy. It is well known that text-based image search suffers from the ambiguity of query keywords. The keywords provided by users tend to be short.

Generally, most photo images stored on the Web have lots of tags added with user's subjective judgments not by the importance of them. So, in tagged Web image retrieval, they have become the cause of precision rate decrease on simple matching of tags to a given query. A common practice to improve search performance is to re-rank 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. In previous works for image search re-ranking suffers from the unreliability of the assumptions under which the initial text-based image search result is employed in the Re-ranking Process. In many cases it is hard for users to describe the visual content of target images using keywords accurately. In order to solve the ambiguity, additional information has to be used to capture users search intention. One way is text based keyword expansion, making the textual description of the query more detailed. Capturing the users search intention from this one-click query image in four steps.

1. The query image is categorized into one of the predefined adaptive weight categories which reflect users search intention at a coarse level. Inside each category, a specific weight schema is used to combine visual features adaptive to this kind of image to better re-rank the text-based search result.
2. Based on the visual content of the query image selected by the user and through image clustering, query keywords are expanded to capture user intention.
3. Expanded keywords are used to enlarge the image pool to contain more relevant images.
4. Expanded keywords are also used to expand the query image to multiple positive visual examples from which new query specific visual and textual similarity metrics are learned to further improve content-based image re-ranking.

II. LITERATURE REVIEW

A. Circular Re-ranking

In [1], Ting Yao proposed Multi-modal graph based and circular re-ranking techniques proposed in recent years capture more than one feature of image for more accurate re-ranking results. The basic idea of circular re-ranking 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. These methods do not always compete but can complement each other.

B. (pLSA) for mining visual categories through clustering of images

Fergus R. et al, [2], 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 re-ranked based on the distance to the mined categories. Self-re-ranking seeks consensus from the initial ranked list as visual patterns for re-ranking.

C. Information Bottleneck (IB) re-ranking

Hsu et al, [4], Smeulders et al, [9] employed information bottleneck (IB) re-ranking to find the clustering of images that preserves the maximal mutual information between the search relevance and visual features. The IB re-ranking 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.

D. Crowd Re-ranking

Richter et al,[10], employed an crowd re-ranking is similar to self-re-ranking 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.

Liu et al. [5] suggested a re-ranking 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 re-ranking. Different from self- and crowd-re-ranking, example-based re-ranking relies on a few query examples provided by users for model learning.

E. Multimedia search with pseudo relevance feedback

Yan et al, [7], Tao Mai et al, [11] 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 re-ranking.

F. Optimizing video search re-ranking via minimum incremental information loss.

Liu et al. [6], 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.

III. SCOPE OF PROPOSED WORK

These current approaches focus on the mining of recurrent patterns from different means, such as by random walk [3], external knowledge [6], and classifier learning [7], web image search engines retrieve and rank images mostly based on surrounding text features.

Image redundancy is still a problem area concerned. It is difficult to interpret users search intention only by query keywords and this leads to ambiguous and noisy search results which are far from satisfactory. It is important to use visual information in order to solve the ambiguity in text-based image retrieval.

A. System Architecture

User first submits query keywords. A pool of images is retrieved by text-based search. Then the user is asked to select a query image from the image pool. The query image is classified as one of the predefined adaptive weight categories. Images in the pool are re-ranked based on their visual similarities to the query image and the similarities are computed using the weight schema specified by the category to combine visual features.

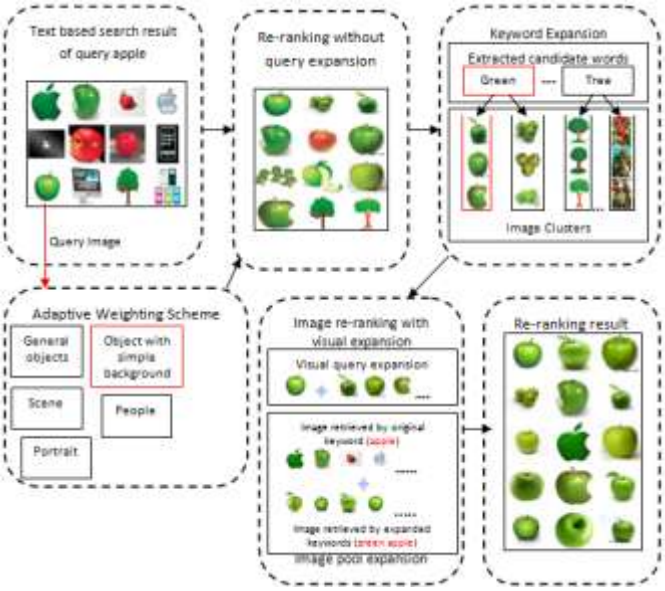

Figure 1: System architecture

In the keyword expansion step, words are extracted from the textual descriptions of the top k images most similar to the query image, and the tf-idf method is used to rank these words and reliable keyword expansions are used further for image clustering. Among all the clusters of different candidate words, cluster with the largest visual similarity to the query image is selected as visual query expansion and its corresponding word is selected to form keyword expansion. A query specific visual similarity metric and a query specific textual similarity metric are learned from both the query image and the visual query expansion. The image pool is enlarged through combining the original image pool retrieved by the query keywords provided by the user and an additional image pool retrieved by the expanded keywords. Images in the enlarged pool are re-ranked using the learned query-specific visual and textual similarity metrics.

B. Adaptive Similarity

How to integrate various visual features to compute the similarities between the query image and other images is an important problem. We have proposed an Adaptive Similarity, motivated by the idea that a user always has specific intention when submitting a query image. In our approach, the query image is firstly categorized into one of the predefined adaptive weight categories. Inside each category, a specific pre-trained weight schema is used to combine visual features adapting to this kind of images to better re-rank the text-based search result. This correspondence between the query image and its proper similarity measurement reflects the user intention.

Algorithm: Feature weight learning for a certain query category

1. Input: Initial weight D_i for all query images i in the current intention category Q_q, similarity matrices s_m $(i,.)$ for all query image i and feature m;
2. Initialize: Set step $t = 1$, set $D_i^1 = D_i$ for all i; while not converged do
   for each query image $i \in Q_q$ do
3. Select best feature m_t and the corresponding similarity s_{mt} (i, **.**) for current re-ranking problem under weight D_i^t ;
4. Calculate ensemble weight $\alpha_{t..}$
5. Adjust weight
   $D_i^{t+1}(j,k) \propto$
   $D_i^t(j,k) \exp\{\propto_t [s_{mt}(i,j) -$
   $smti, k;$
6. Normalize D_i^{t+1} to make it a distribution;
7. t++;
   end for
   end while
8. Output: Final optimal similarity measure for current intention category:

$$s^q(.,.) = \frac{\sum_t \propto_t s_{m_t}(.,.)}{\sum_t \propto_t},$$

And the weight for feature m: $a_m^q = \frac{\sum_{m_{t=m}} \propto_t}{\sum_t \propto_t}$

C. Keyword expansion

Query keywords given by users tend to be short and some important keywords may be missed because of user's lack of knowledge on the textual description of target images. In our approach, query keywords are expanded to capture users search intention. The expanded keywords better capture users search intention since the consistency of both visual content and textual description is ensured.

Once the top k images most similar to the query image are found according to the visual similarity metric introduced in adaptive similarity, words from their textual descriptions are extracted and ranked, using the *term frequency-inverse document frequency* (tf-idf) [13] method.

We do keyword expansion through image clustering. For each candidate word w_i, all the images containing w_i in the image pool are found. However, they cannot be directly used as the visual representations of w_i for two reasons. First, there may be a number of noisy images irrelevant to w_i. Second, even if these images are relevant to w_i semantically, they may have quite different visual content. In order to find images with similar visual content as the query example and remove noisy images, we divide these images into different clusters using k-Means.

In order to find images with similar visual content as the query example and remove noisy images, we divide these images into different clusters using k-Means. The number of clusters is empirically set to be $n/6$ where n is the number of images to cluster.

Each word w_i has t_i clusters $C(w_i) = \{c_{i,1}, \ldots\ldots, c_{i,t_i}\}$. The visual distance between the query image and a cluster c is calculated as the mean of the distances between the query image and the images in c. The cluster $c_{i,j}$ with the minimal distance is chosen as visual query expansion and its corresponding word w_i, combined with the original keyword query q, is chosen as keyword expansion q'.

If the distance between the closest cluster and the query image is larger than a threshold ρ, it indicates that there is no suitable image cluster and word to expand the query, and thus query expansion will not be used.

D. Visual Query Expansion

The goal of visual query expansion is to obtain multiple positive example images to learn a visual similarity metric which is more robust and more specific to the query image. By adding more positive examples to learn a more robust similarity metric, irrelevant images can be filtered out. The clusters of images chosen in keyword expansion have the closest visual distance to the query example and have consistent semantic meanings. Thus they are used as additional positive examples for visual query expansion. We adopt the one-class SVM [14] to refine the visual similarity. The one-class SVM classifier is trained from the additional positive examples obtained by visual query expansion. An image to be re-ranked is input to the one-class SVM classifier and the output is used as the similarity to the query image.

The one-class SVM classifier is trained from the additional positive examples obtained by visual query expansion. It requires defining the kernel between images, and the kernel is computed from the similarity.

Kernel methods can be thought of as instance-based learners rather than learning some fixed set of parameters corresponding to the features of their inputs, they instead remember the i^{th} training example (x_i, y_i) by learning a corresponding weight (w_i). Prediction for unlabeled inputs, i.e., those not in the training set, is treated by the application of a similarity function k, called a kernel, between the unlabeled input x' and each of the training inputs x_i. For instance, a kernelized binary classifier typically computes a weighted sum of similarities

$$y = sgn \sum_{i=1}^{n} w_i y_i k(x_i, x'),$$

Where $y \in \{-1, +1\}$ is the kernelized binary classifier's predicted label for the unlabeled input x' whose hidden true label y is of interest.

$k: X \times X \rightarrow \mathbb{R}$ is the kernel function that measures similarity between any pair of inputs $x \in X$ and $x' \in X$ and,

$D = \{(x_i, y_i)\}_{i=1}^{n}$ are the n labelled examples in the classifier's training set D, where the training labels $y_i \in \{-1, +1\}$, and the $w_i \in \mathbb{R}$ coefficients are the weights for the training examples.

The sign function sgn determines whether the predicted classification y comes out positive or negative.

E. Image pool expansion

The image pool retrieved by text-based search accommodates images with a large variety of semantic meanings and the number of images related to the query image is small. In this case, re-ranking images in the pool is not very effective. Thus more accurate query by keywords is needed to narrow the intention and retrieve more relevant images. A naive way is to ask the user to click on one of the suggested keywords given by traditional approaches only using text information and to expand query results. This increases user's burden. Moreover, the suggested keywords based on text information only are not accurate to describe user's intention. Keyword expansions suggested by our approach using both visual and textual information better capture user's intention. They are automatically added into the text query and enlarge the image pool to include more relevant images.

Considering efficiency, image search engines such as Bing image search only re-rank the top N images of the text-based image search result. If the query keywords do not capture the user's search intention accurately, there are only a small number of relevant images with the same semantic meanings as the query image in the image pool. This can significantly degrade the ranking performance. In adaptive similarity, we re-rank the top N retrieved images by the original keyword query based on their visual similarities to the query image. We remove the N/2 images with the lowest ranks from the image pool. Using the expanded keywords as query, the top N/2 retrieved images are added to the image pool. We believe that there are more relevant images in the image pool with the help of expanded query keywords.

F. Combining Visual and Textual Similarities

Learning a query specific textual similarity metric from the positive examples obtained by visual query expansion and combining it with the query specific visual similarity metric introduced in visual query expansion can further improve the performance of image re-ranking and we will get re-ranked result.

REFERENCES

[1] Ting Yao, Chong-Wah Ngo and Tao Mei, "Circular Reranking for Visual Search", IEEE 2013
[2] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning objects categories from Google's image search," in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2005, pp. 1816–1823.
[3] 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.
[4] 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.
[5] Y. Liu, T. Mei, and X.-S.Hua, "Crowd Reranking: Exploring multiple search engines for visual search reranking," in Proc. ACM Special Interest Group Inf. Retr., 2009, pp. 500–507.
[6] 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.
[7] R. Yan, A. Hauptmann, and R. Jin, "Multimedia search with pseudo relevance feedback," in Proc. ACM Int. Conf. Image Video Retr., 2003, pp. 238–247.
[8] Bing. (2009) [Online]. Available: http://www.bing.com
[9] A. W. M. Smeulders, M.Worring, S. Santini, and A. Gupta, Content based image retrieval at the end of the early years, IEEE Trans Pattern Anal. Mach Intel., vol. 22, no. 12, pp. 1349–1380, Dec. 2000.
[10]F. Richter, S. Romberg, E. Horster, and R. Lienhart, "Multimodal ranking for image search on community databases," in Proc. ACM SIGMM Int. Workshop Multimedia Inf. Retr. , 2010, pp. 63–72.
[11]Tao Mai, Yong Rui, Shipeng Li, Microsoft Research Asia, Qi Tian, Multimedia Search Reranking: A literature Survey, ACM Journal, Vol. 2, No., 3, Article 1, May 2012.
[12]Xiaopeng Yang, Yongdong Zhang, Ting Yao,Chong-Wah Ngo, Tao Mei; Click-boosting multi-modality graph-based reranking for image Search, Multimedia Systems, DOI 10.1007/s00530-014-0379-8, Springer- Verlag Berlin Heidelberg 2014, Published online 09 May 2014.
[13] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. Addison-Wesley Longman Publishing Co., 1999.
[14]Y. Chen, X. Zhou, and T. Huang, "One-Class SVM for Learning in Image Retrieval," Proc. IEEE Int'l Conf. Image Processing, 2001.