Volume 5, Issue 3, March 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Image Re-Ranking Using Query-specific Semantic Signatures

[1]Tejashree Kumar Shinde*, [2]Prof. Prakash. B. Dhainje, [3]Dr. Deshmukh Pradeep K.
[1]M.E (CSE) Student SIETC, Paniv CSE Dept., Solapur University, Solapur, Maharashtra, India
[2]MTech, MBA (IT), PhD (CSE) Head of CSE Dept. SIETC, Paniv CSE Dept., Solapur University, Solapur, M.H., India
[3]M.E., PhD (CSE) Principal, SIETC, Paniv CSE Dept., Solapur University, Solapur, Maharashtra, India

Abstract- In this paper, we propose a novel image re-ranking framework, which automatically offline learns different visual semantic spaces for different query keywords through keyword expansions. The visual features of images are projected into their related visual semantic spaces to get semantic signatures. At the online stage, images are re-ranked by comparing their semantic signatures obtained from the visual semantic space specified by the query keyword. The new approach significantly improves both the accuracy and efficiency of image re-ranking. The original visual features of thousands of dimensions can be projected to the semantic signatures as short as 25 dimensions.

Keywords— semantic, re-ranking, query, search engine, framework.

I. INTRODUCTION

Web-scale image search engines mostly use keywords as queries and rely on surrounding text to search images. It is well known that they suffer from the ambiguity of query keywords. For example, using "apple" as query, the retrieved images belong to different categories, such as "red apple", "apple logo", and "apple laptop". Online image reranking has been shown to be an effective way to improve the image search results [5, 4, 9]. Major internet image search engines have since adopted the re-ranking strategy [5]. Its diagram is shown in Figure 1. Given a query keyword input by a user, according to a stored word-image index file, a pool of images relevant to the query keyword are retrieved by the search engine. By asking a user to select a query image, which reflects the user's search intention, from the pool, the remaining images in the pool are re-ranked based on their visual similarities with the query image. The visual features of images are pre-computed offline and stored by the search engine. The main online computational cost of image re-ranking is on comparing visual features. In order to achieve high efficiency, the visual feature vectors need to be short and their matching needs to be fast. Another major challenge is that the similarities of low level visual features may not well correlate with images' high-level semantic meanings which interpret users' search intention. To narrow down this semantic gap, for offline image recognition and retrieval, there have been a number of studies to map visual features to a

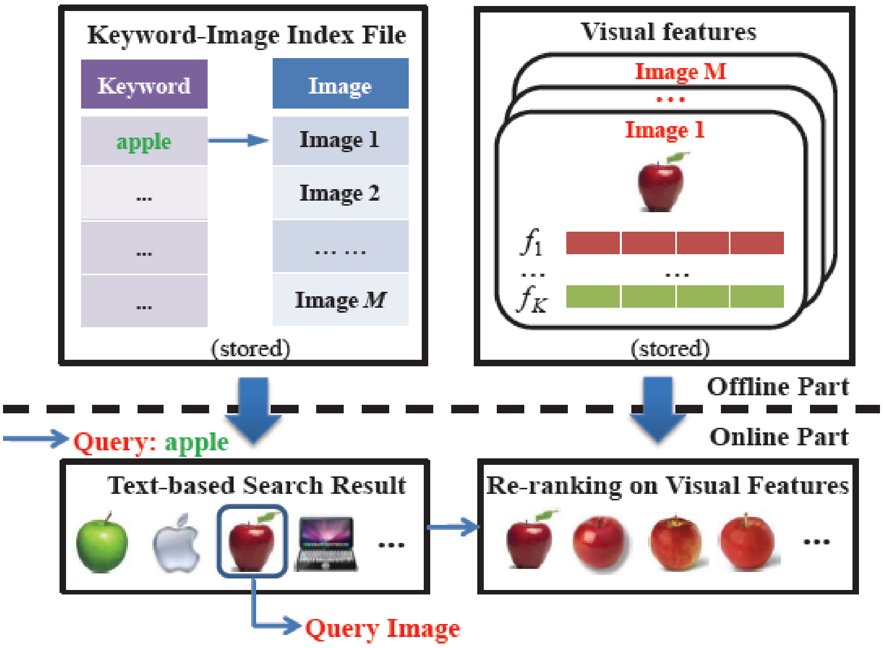

Figure 1. The conventional image re-ranking framework.

set of predefined concepts or attributes as semantic signature [11, 7, 15]. However, these approaches are only applicable to closed image sets of relatively small sizes. They are not suitable for online web- based image re-ranking.

A. Approach

In this paper, a novel framework is proposed for web image re-ranking. Instead of constructing a universal concept dictionary, it learns different visual semantic spaces for different query keywords individually and automatically. We believe that the semantic space related to the images to be re-ranked can be significantly narrowed down by the query keyword provided by the user. For example, if the query keyword is "apple", the semantic concepts of "mountains" and "Paris" are unlikely to be relevant and can be ignored. Instead, the semantic concepts of "computers" and "fruit" will be used to learn the visual semantic space related to "apple". The query-specific visual semantic spaces can more accurately model the images to be re-ranked, since they have removed other potentially unlimited number of non-relevant concepts, which serve only as noise and deteriorate the performance of re-ranking in terms of both accuracy and computational cost. The visual features of images are then projected into their related visual semantic spaces to get semantic signatures. At the online stage, images are re-ranked by comparing their semantic signatures obtained.

B. Related Work

Content-based image retrieval uses visual features to calculate image similarity. Relevance feedback [13, 16, 14] was widely used to learn visual similarity metrics to capture users' search intention. However, it required more users' effort to select multiple relevant and irrelevant image examples and often needs online training. For a web-scale commercial system, users' feedback has to be limited to the minimum with no online training. Cui et al. [5, 4] proposed an image re-ranking approach which limited users' effort to just one-click feedback. Such simple image re-ranking approach has been adopted by popular web-scale image search engines such as Bing and Google recently, as the "find similar images" function.

The key component of image re-ranking is to compute the visual similarities between images. Many image features [8, 6, 2, 10] have been developed in recent years. However, for different query images, low-level visual features that are effective for one image category may not work well for another. To address this, Cui et al. [5, 4] classified the query images into eight predefined intention categories and gave different feature weighting schemes to different types of query images. However, it was difficult for only eight weighting schemes to cover the large diversity of all the web images. It was also likely for a query image to be classified to a wrong category.

Recently, for general image recognition and matching, there have been a number of works on using predefined concepts or attributes as image signature. Rasiwasia et al. [11] mapped visual features to a universal concept dictionary. Lampert et al. [7] used predefined attributes with semantic meanings to detect novel object classes. Some approaches [1, 15, 12] transferred knowledge between object classes by measuring the similarities between novel object classes and known object classes (called reference classes). All these concepts/attributes/reference-classes were universally applied to all the images and their training data was manually selected. They are more suitable for offline databases with lower diversity (such as animal databases [7, 12] and face databases [15]) such that object classes better share similarities. To model all the web images, a huge set of concepts or reference classes are required, which is impractical and ineffective for online image re-ranking.

II. APPROACH OVERVIEW

The diagram of our approach is shown in Figure 2. At the offline stage, the reference classes (which represent different semantic concepts) of query keywords are automatically discovered. For a query keyword (e.g. "apple"), a set of most relevant keyword expansions (such as "red apple", "apple macbook", and "apple iphone") are automatically selected considering both textual and visual information. This set of keyword expansions defines the reference classes for the query keyword. In order to automatically obtain the training examples of a reference class, the keyword expansion (e.g. "red apple") is used to retrieve images by the search engine. Images retrieved by the keyword expansion ("red apple") are much less diverse than those retrieved by the original keyword ("apple"). After automatically removing outliers, the retrieved top images are used as the training examples of the reference class. Some reference classes (such as "apple laptop" and "apple macbook") have similar semantic meanings and their training sets are visually similar. In order to improve the efficiency of online image re-ranking, redundant reference classes are removed. For each query keyword, a multi-class classifier on low level visual features is trained from the training sets of its reference classes and stored offline. If there are K types of visual features, one could combine them to train a single classifier. It is also possible to train a separate classifier for each type of features. Our experiments show that the latter choice can increase the re-ranking accuracy but will also increase storage and reduce the online matching efficiency because of the increased size of semantic signatures.

An image may be relevant to multiple query keywords. Therefore it could have several semantic signatures obtained in different semantic spaces. According to the word image index file, each image in the database is associated with a few relevant keywords. For each relevant keyword, a semantic signature of the image is extracted by computing the visual similarities between the image and the reference classes of the keyword using the classifiers trained in the previous step. The reference classes form the basis of the semantic space of the keyword. If an image has N relevant keywords, then it has N semantic signatures to be computed and stored offline. At the online stage, a pool of images are retrieved by the search engine according to the query keyword input by a user. Since all the images in the pool are relevant to the query keyword, they all have pre-computed semantic signatures in the semantic space of the query keyword. Once the user chooses a query image, all the images are re-ranked by comparing similarities of the semantic signatures.

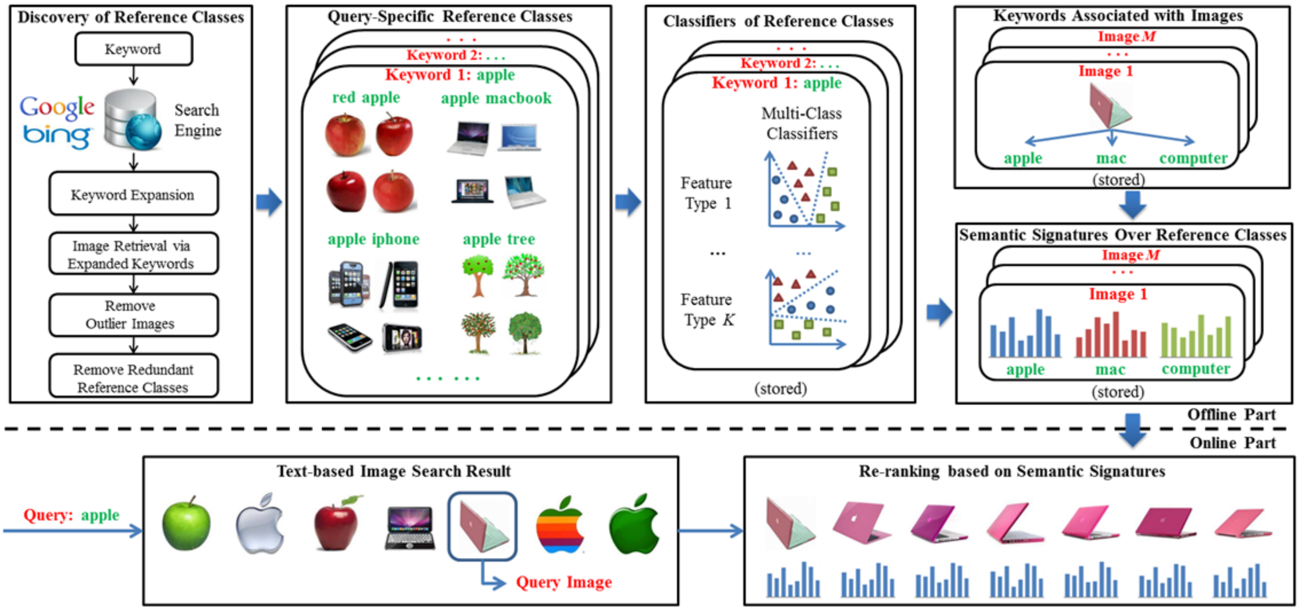

Figure 2. Diagram of our new image re-ranking framework

III. SEMANTIC SIGNATURES

Given M reference classes for keyword q and their training images automatically retrieved, a multi-class classifier on the visual features of images is trained and it outputs an M-dimensional vector p, indicating the probabilities of a new image I belonging to different reference classes. Then p is used as semantic signature of I. The distance between two images Ia and Ib are measured as the L1-distance between their semantic signatures pa and pb,

$$d(Ia;\ Ib) = ||p^{a}–p^{b}\ ||_{1}$$

A. *Combined Features vs Separate Features*

In order to train the SVM classifier, we adopt six types of visual features used in [5]: attention guided color signature, color spatialet, wavelet, multi-layer rotation invariant edge orientation histogram, histogram of gradients, and GIST. They characterize images from different perspectives of color, shape, and texture. The combined features have around 1; 700 dimensions in total. A natural idea is to combine all types of visual features to train a single powerful SVM classifier which better distinguish different reference classes. However, the purpose of using semantic signatures is to capture the visual content of an image, which may belong to none of the reference classes, instead of classifying it into one of the reference classes. If there are N types of independent visual features, it is actually more effective to train separate SVM classifiers on different types of features and to combine the N semantic signatures fpngN n=1 from the outputs of N classifiers. The N semantic signatures describe the visual content of an image from different aspects (e.g. color, texture, and shape) and can better characterize images outside the reference classes. For example, in Figure 3, “red apple” and “apple tree” are two reference classes. A new image of “green apple” can be well characterized by two semantic signatures from two classifiers trained on color features and shape features separately, since “green apple” is similar to “red apple” in shape and similar to “apple tree” in color. Then the distance between two images Ia and Ib is,

$$d(Ia;\ Ib) = \sum_{n=1}^{N} w_{n}\ ||p^{a,n}–p^{b,n}\ ||_{1}$$

where w_n is the weight on different semantic signatures and it is specified by the query image Ia selected by the user. Wn

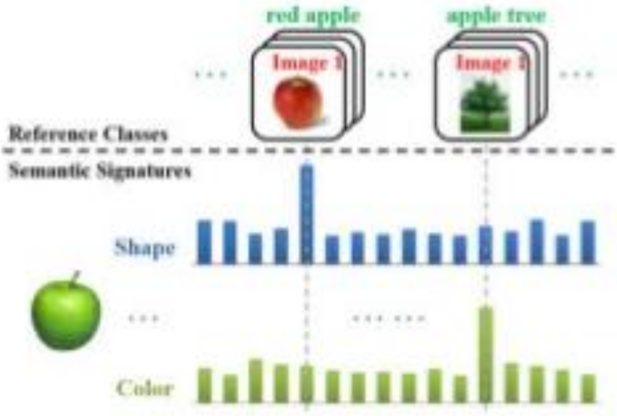

Figure 3. Describe “green apple” using reference classes. Its shape is captured by shape classifier of “red apple“ and its color is captured by color classifier of “apple tree”.

is decided by the entropy of pa;n,

$$wn = \frac{1}{1 + eH(pa;n)} ;$$

$$H(pa;n) =- \sum_{i=1}^{M} P_i^{a,n} \ln P_i^{a,n}$$

If pa;n uniformly distributes over reference classes, the nth type of visual features of the query image cannot be well characterized by any of the reference classes and we assign a low weight to this semantic signature.

IV. RERANKING PRECISIONS

Averaged top m precision is used as the evaluation criterion. Top m precision is defined as the proportion of relevant images among top m re-ranked images. Relevant images are those in the same category as the query image. Averaged top m precision is obtained by averaging top m precision for every query image (excluding outliers). We adopt this criterion instead of the precision-recall curve since in image re-ranking, the users are more concerned about the qualities of top retrieved images instead of number of relevant images returned in the whole result set. We compare with two benchmark image re-ranking approaches used in [5]. They directly compare visual features. (1) Global Weighting. Predefined fixed weights are adopted to fuse the distances of different low-level visual features. (2) Adaptive Weighting. [5] proposed adaptive weights for query images to fuse the distances of different low-level visual features. It is adopted by Bing Image Search. For our new approaches, two different ways of computing semantic signatures as discussed in Section 4.1 are compared.

_ Query-specific visual semantic space using single signatures (QSVSS Single). For an image, a single semantic signature is computed from one SVM classifier trained by combining all types of visual features.

_ Query-specific visual semantic space using multiple signatures (QSVSS Multiple). For an image, multiple semantic signatures are computed from multiple SVM classifiers, each of which is trained on one type of visual features separately. Some parameters used in our approach as mentioned in Sec-tions 3 and 4 are tuned in a small separate data set and they are fixed in all the experiments. Our approach significantly outperforms GlobalWeighting and AdaptiveWeighting, which directly compare visual features. On data set I, our approach enhances the averaged top 10 precision from 44:41% (Adaptive Weighting) to 55:12% (QSVSS Multiple). 24:1% relative improvement has been achieved. In our approach, computing multiple semantic signatures from separate visual features has higher precisions than computing a single semantic signature from combined features. However, it costs more online computation since the dimensionality of multiple semantic signatures is higher. if the testing images for re-ranking and images of reference classes are collected from different search engines, the performance is slightly lower than the case when they are collected from the same search engine. However, it is still much higher than directly comparing visual features. This indicates that we can utilize images from various sources to learn query-specific semantic spaces. even if the testing images and images of reference classes are collected at different times (eleven months apart), query specific semantic spaces still can effectively improve re-ranking. Compared with Adaptive Weighting, the averaged top 10 precision has been improved by 6:6% and the averaged top 100 precision has been improved by 9:3%. This indicates that once the query-specific semantic spaces are learned, they can remain effective for a long time and do not have to be updated frequently.

A. Reranking

Images outside the reference classes It is interesting to know whether the learned queryspecific semantic spaces are effective for query images which are outside the reference classes. To answer this question, if the category of an query image corresponds to a reference class, we deliberately delete this reference class and use the remaining reference classes to train SVM classifiers and to compute semantic signatures when comparing this query image with other images. We repeat this for every image and calculate the average top m precisions. This evaluation is denoted as RmCategoryRef and is done on data set III6. Multiple semantic signatures (QSVSS Multiple) are used. The results are shown in Figure 5. It still greatly outperforms the approaches of directly comparing visual features. This result can be explained from two aspects. (1) As discussed in Section 4.1, the multiple semantic signatures obtained from different types of visual features separately have the capability to characterize the visual content of images outside the reference classes. (2) Many negative examples (images belonging to different categories than the query image) are well modeled by the reference classes and are therefore pushed backward on the ranking list.

B. Query specific semantic space vs. universal semantic space

In previous works [11, 7, 1, 15, 12], a universal set of reference classes or concepts were used to map visual features to a semantic space for object recognition or image retrieval on closed databases. In this experiment, we evaluate whether this approach is applicable to web-based image re-ranking and compare it with our approach. We randomly select M reference classes from the whole set of reference classes of all the 120 query keywords in data set I. The M selected reference classes are used to train a universal semantic space in a way similar to Section 4.1. Multiple semantic signatures are obtained from different types of features separately. This universal semantic space is applied to data set III for image re-ranking.. M is chosen as 25, 80, 120 and 1607. This method is denoted as UnivMClasses. When the universal semantic space chooses the same number (25) of reference classes as our query-specific semantic spaces, its precisions are no better than visual features. Its precisions increase when a larger number of reference classes are selected. However, the gain increases very slowly when M is larger than 80. Its best precisions (when M = 160) are much lower than QSVSS Multiple and even lower than RmCategoryRef, even though the length of its semantic signatures is five times larger than ours.

C. User study

User experience is critical for web-based image search. In order to fully reflect the extent of users' satisfaction, user study is conducted to compare the results of our approach (QSVSS Multiple) compared with Adaptive Weighting on data set I. Twenty users are invited. Eight of them are familiar with image search and the other twelve are not. To avoid bias on the evaluation, we ensure that all the participants do not have any knowledge about the current approaches for image re-ranking, and they are not told which results are from which methods. Each user is assigned 20 queries and is asked to randomly select 30 images per query. Each selected image is used as a query image and the re-ranking results of Adaptive Weighting and our approach are shown to the user. The user is required to indicate whether our re-ranking result is "Much Better", "Better", "Similar", "Worse", or "Much Worse" than that of Adaptive Weighting. 12; 000 user comparison results are collected. The comparison results are shown in Figure 6. In over 55% cases our approach delivers better results than Adaptive Weighting and only in less than 18% cases ours is worse, which are often the noisy cases with few images relevant to the query image exists.

V. CONCLUSIONS

We propose a novel image re-ranking framework, which learns query-specific semantic spaces to significantly improve the effectiveness and efficiency of online image reranking. The visual features of images are projected into their related visual semantic spaces automatically learned through keyword expansions at the offline stage. The extracted semantic signatures can be 70 times shorter than the original visual feature on average, while achieve 20%□35% relative improvement on re-ranking precisions over state-ofthe- art methods.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Dr. Deshmukh Pradeep K for their guidance and to Prof. Prakash. B. Dhainje for their supervision . We would like to express our appreciation to our parents and all the teachers and lecturers who help us to understand the importance of knowledge and show us the best way to gain it.

REFERENCES

[1] E. Bart and S. Ullman. Single-example learning of novel classes using representation by similarity. In Proc. BMVC, 2005.

[2] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang. Spatial-bag-offeatures. In Proc. CVPR, 2010.

[3] G. Cauwenberghs and T. Poggio. Incremental and decremental support vector machine learning. In Proc. NIPS, 2001.

[4] J. Cui, F. Wen, and X. Tang. Intentsearch: Interactive on-line image search re-ranking. In Proc. ACM Multimedia. ACM, 2008.

[5] J. Cui, F. Wen, and X. Tang. Real time google and live image search re-ranking. In Proc. ACM Multimedia, 2008.

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Proc. CVPR, 2005.

[7] C. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In Proc. CVPR, 2005.

[8] D. Lowe. Distinctive image features from scale-invariant keypoints. Int'l Journal of Computer Vision, 2004.

[9] B. Luo, X. Wang, and X. Tang. A world wide web based image search engine using text and image content features. In Proceedings of the SPIE Electronic Imaging, 2003.

[10] J. Philbin, M. Isard, J. Sivic, and A. Zisserman. Descriptor Learning for Efficient Retrieval. In Proc. ECCV, 2010.

[11] N. Rasiwasia, P. J. Moreno, and N. Vasconcelos. Bridging the gap: Query by semantic example. IEEE Trans. on Multimedia, 2007.

[12] M. Rohrbach, M. Stark, G. Szarvas, I. Gurevych, and B. Schiele What helps wherevand why? semantic relatedness for knowledge transfer. In Proc. CVPR, 2010.

[13] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: a power tool for interactive content-based image retrieval. IEEE Trans. on Circuits and Systems for Video Technology, 1998.

[14] D. Tao, X. Tang, X. Li, and X. Wu. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006.

[15] Q. Yin, X. Tang, and J. Sun. An associate-predict model for face recognition. In Proc. CVPR, 2011.

[16] X. S. Zhou and T. S. Huang. Relevance feedback in image retrieval: A comprehensive review. Multimedia Systems, 2003.