International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Web-based Image Search using Query-Specific Semantic Signatures

Dipak Pardhi[1], Prajakta Raghunath Pachpande[2]

[1, 2]Godavari Foundation's Godavari College of Engineering, North Maharashtra University, Jalgaon 425001, India

Abstract: ***Image re-ranking is a useful method for web-based image search. The search based on only keywords queried by the users is not efficient and results in imprecise output. The web-based image search used by Bing and Google uses image re-ranking. In image re-ranking, users' intention is captured by one-click on the query image. This helps in providing better search results to the users. In this paper, we review the method in which a query keyword is first used to retrieve a plethora of images based on the keyword. Image re-ranking framework automatically learns different semantic spaces offline for different query keywords. To get semantic signatures for images, their visual features are projected into their related semantic spaces. Images are re-ranked by comparing their semantic signatures and the query keyword during the online stage. The query-specific semantic signatures, in the reviewed paper, significantly improve both the accuracy and efficiency of the re-ranking process. Hence, it is proved to be a better method than the conventional web-based image search techniques.***

Keywords: Image search, image re-ranking, semantic space, semantic signature, keyword expansion.

1. Introduction

WEB-SCALE image search engines mostly use keywords as queries and rely on surrounding text to search images. They suffer from the ambiguity of query keywords, because it is hard for users to accurately describe the visual content of target images only using keywords. For example, using "apple" as a query keyword, the retrieved images belong to different categories (also called concepts in this paper), such as "red apple," "apple logo," and "apple laptop." In order to solve the ambiguity, content- based image retrieval [1], [2] with relevance feedback [3], [4], [5] is widely used. It requires users to select multiple relevant and irrelevant image examples, from which visual similarity metrics are learned through online training. Images are re-ranked based on the learned visual similarities. However, for web-scale commercial systems, users' feedback has to be limited to the minimum without online training.

Online image re-ranking [6], [7], [8], which limits users' effort to just one-click feedback, is an effective way to improve search results and its interaction is simple enough. Major web image search engines have adopted this strategy [8]. Its diagram is shown in Fig. 1. Given a query keyword input by a user, a pool of images relevant to the query keyword is retrieved by the search engine according to a stored word-image index file. Usually the size of the returned image pool is fixed, e.g., containing 1000 images. By asking the 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 word image index file and visual features of images are precomputed offline and stored. The main online computational cost is on comparing visual features. To achieve high efficiency, the visual feature vectors need to be short and their matching needs to be fast. Some popular visual features are in high dimensions and efficiency is not satisfactory if they are directly matched.

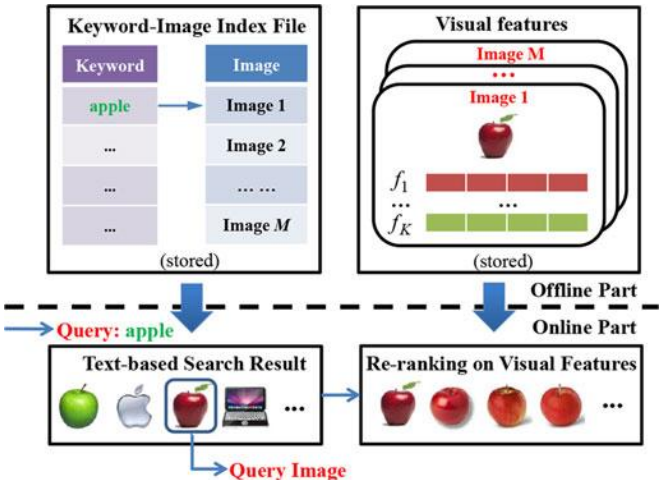

Figure 1: The conventional image re-ranking framework.

Another major challenge is that, without online training, the similarities of low-level visual features may not well correlate with images' high-level semantic meanings which interpret users' search intention. Some examples are shown in Fig. 2. Moreover, low-level features are sometimes inconsistent with visual perception. For example, if images of the same object are captured from different viewpoints, under different lightings or even with different compression artifacts, their low-level features may change significantly, although humans think the visual content does not change much. To reduce this semantic gap and inconsistency with visual perception, there have been a number of studies to map visual features to a set of predefined concepts or attributes as semantic signatures [9], [10], [11], [12]. For example, Kovashka et al. [12] proposed a system which refined image search with relative attribute feedback. Users described their search intention with reference images and a set of pre-defined attributes. These concepts and attributes are pre-trained offline and have tolerance with variation of visual content. However, these approaches are only applicable to closed image sets of relatively small sizes, but not suitable for online web-scale image re-ranking. According to our empirical study, images retrieved by 120

query keywords alone include more than 1,500 concepts. It is difficult and inefficient to design a huge concept dictionary to characterize highly diverse web images. Since the topics of web images change dynamically, it is desirable that the concepts and attributes can be automatically found instead of being manually defined.

Figure 2: All the images shown in this figure are related to palm trees. They are different in color, shape, and texture.

2. Related Work

The key component of image re-ranking is to compute visual similarities reflecting semantic relevance of images. Many visual features have been developed in recent years. However, for different query images, the effective low-level visual features are different. Therefore, Cui et al. [6], [7] classified query images into eight predefined intention categories and gave different feature weighting schemes to different types of query images. But it was difficult for the 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. In order to reduce the semantic gap, query-specific semantic signature was first proposed in Kuo et al. recently augmented each image with relevant semantic features through propagation over a visual graph and a textual graph which were correlated.

Another way of learning visual similarities without adding users' burden is pseudo relevance feedback. It takes the top N images most visually similar to the query image as expanded positive examples to learn a similarity metric. Since the top N images are not necessarily semantically-consistent with the query image, the learned similarity metric may not reliably reflect the semantic relevance and may even deteriorate re-ranking performance. In object retrieval, in order to purify the expanded positive examples, the spatial configurations of local visual features are verified. But it is not applicable to general web image search, where relevant images may not contain the same objects.

Recently, for general image recognition and matching, there have been a number of works on using projections over predefined concepts, attributes or reference classes as image signatures. The classifiers of concepts, attributes, and reference classes are trained from known classes with labeled examples. But the knowledge learned from the known classes can be transferred to recognize samples of novel classes which have few or even no training samples. Since these concepts, attributes, and reference classes are defined with semantic meanings, the projections over them can well capture the semantic meanings of new images even without further training. Rasiwasia et al. [9] mapped visual features to a universal concept dictionary for image retrieval. Attributes with semantic meanings were used for object detection [10] object recognition face recognition image search action recognition and 3D object retrieval. Lampert et al. [10] predefined a set of attributes on an animal database and detected target objects based on a combination of human-specified attributes instead of training images. Sharmanska et al. augmented this representation with additional dimensions and allowed a smooth transition between zero-shot learning, unsupervised training and supervised training. Parikh and Grauman proposed relative attributes to indicate the strength of an attribute in an image with respect to other images. Parkash and Parikh used attributes to guide active learning. In order to detect objects of many categories or even unseen categories, instead of building a new detector for each category, Farhadi et al. learned part and attribute detectors which were shared across categories and modeled the correlation among attributes. Some approaches [11] transferred knowledge between object classes by measuring the similarities between novel object classes and known object classes (called reference classes). For example, Torresani et al. proposed an image descriptor which was the output of a number of classifiers on a set of known image classes, and used it to match images of other unrelated visual classes. In the current approaches, all the concepts / attributes / reference-classes are universally applied to all the images and they are manually defined. They are more suitable for offline databases with lower diversity (such as animal databases [10] and face databases [11]), since image classes in these databases can 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. Intuitively, only a small subset of the concepts is relevant to a specific query. Many concepts irrelevant to the query not only increase the computational cost but also deteriorate the accuracy of reranking. However, how to automatically find such relevant concepts and use them for online web image re-ranking was not well explored in previous studies.

3. Approach Overview

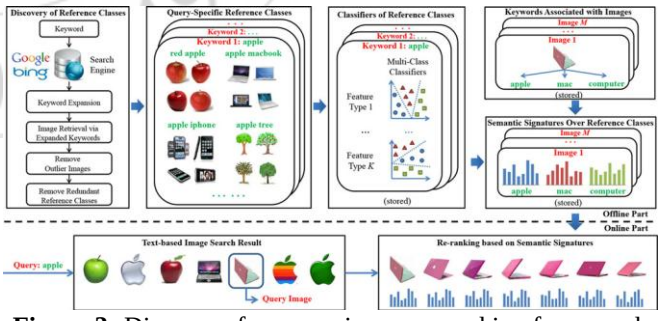

Figure 3: Diagram of our new image re-ranking framework

The diagram of our approach is shown in Fig. 3. It has offline and online parts. At the offline stage, the reference classes (which represent different concepts) related to query keywords are automatically discovered and their training images are automatically collected in several steps. For a query keyword (e.g., "apple"), a set of most relevant keyword expansions (such as "red apple" and "apple MacBook") are automatically selected utilizing 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 based on textual information again. 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. To better measure the similarity of semantic signatures, the semantic correlation between reference classes is estimated with a web-based kernel function.

For each query keyword, its reference classes forms the basis of its semantic space. A multi-class classifier on visual and textual features is trained from the training sets of its reference classes and stored offline. Under a query keyword, the semantic signature of an image is extracted by computing the similarities between the image and the reference classes of the query keyword using the trained multiclass classifier. If there are K types of visual/textual features, such as color, texture, and shape, one could combine them together to train a single classifier, which extracts one semantic signature for an image. It is also possible to train a separate classifier for each type of features. Then, the K classifiers based on different types of features extract K semantic signatures, which are combined at the later stage of image matching. Our experiments show that the latter strategy can increase the re-ranking accuracy at the cost of storage and online matching efficiency because of the increased size of semantic signatures.

According to the word-image index file, an image may be associated with multiple query keywords, which have different semantic spaces. Therefore, it may have different semantic signatures. The query keyword input by the user decides which semantic signature to choose. As an example shown in Fig. 3, an image is associated with three keywords "apple," "mac" and "computer." When using any of the three keywords as query, this image will be retrieved and re-ranked. However, under different query keywords, different semantic spaces are used. Therefore an image could have several semantic signatures obtained in different semantic spaces. They all need to be computed and stored offline.

At the online stage, a pool of images is retrieved by the search engine according to the query keyword. Since all the images in the pool are associated with the query keyword according to the word-image index file, they all have pre-computed semantic signatures in the same semantic space specified by the query keyword. Once the user chooses a query image, these semantic signatures are used to compute image similarities for re-ranking. The semantic correlation of reference classes is incorporated when computing the similarities.

4. Discovery of Reference Classes

4.1 Keyword Expansion

An intuitive way of finding keyword expansions could be first clustering images with visual/textual features and then finding the most frequent word in each cluster as the keyword expansion. We do not adopt this approach for two reasons. Images belonging to the same semantic concept (e.g., "apple laptop") have certain visual diversity (e.g., due to variations of viewpoints and colors of laptops). Therefore, one keyword expansion falls into several image clusters. Similarly, one image cluster may have several keyword expansions with high frequency, because some concepts have overlaps on images. For examples, an image may belong to "Paris Eiffel tower," "Paris nights" and "Paris Album." Since the one-to-one mapping between clusters and keyword expansions do not exist, a post processing step similar to our approach is needed to compute the scores of keywords selected from multiple clusters and fuse them. The multimodal and overlapping distributions of concepts can be well handled by our approach. Secondly, clustering web images with visual and textual features is not an easy task especially with the existence of many outliers. Bad clustering result greatly affects later steps. Since we only need keyword expansions, clustering is avoided in our approach. For each image I, our approach only considers its D nearest neighbors and is robust to outliers.

4.2 Training Images of Reference Classes

In order to automatically obtain the training images of reference classes, each keyword expansion *e* combined with the original keyword *q* is used as query to retrieve images from the search engine and top *K* images are kept. Since the expanded keywords *e* has less semantic ambiguity than the original keyword *q*, the images retrieved by *e* are much less diverse. After removing outliers by k-means clustering, these images are used as the training examples of the reference class. The cluster number of k-means is set as 20 and clusters of sizes smaller than 5 are removed as outliers.

5. Semantic Signatures

In figure 4, "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. If the color and shape features are combined to compute a single semantic signature, it cannot well characterize the image of "green apple." Since the "green apple" is dissimilar to any reference class when jointly considering color and shape, the semantic signature has low distributions over all the reference classes.

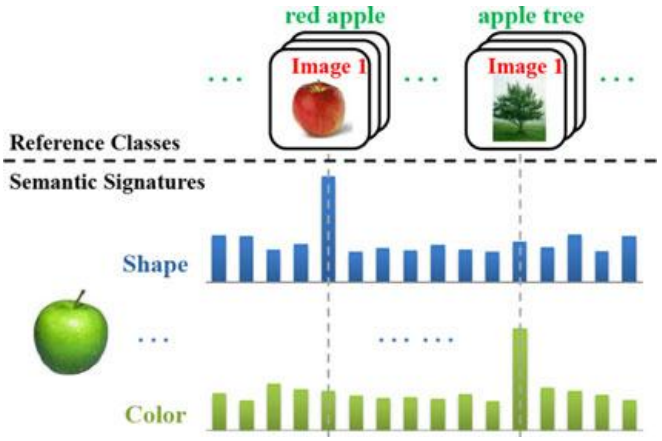

Figure 4: Describe "green apple" with reference classes. Its shape is captured by the shape classifier of "red apple" and its color is captured by the color classifier of "apple tree."

6. Experimental Results

The images for testing the performance of re-ranking and the training images of reference classes can be collected at different time (since the update of reference classes may be delayed) and from different search engines. Given a query Keyword, 1,000 images are retrieved from the whole web using a search engine. As summarized in Table 1, we create three data sets to evaluate the performance of our approach in different scenarios. In data set I, 120; 000 testing images for re-ranking were collected from the Bing Image Search with 120 query keywords in July 2010. These query keywords cover diverse topics including animals, plants, food, places, people, events, objects, and scenes, etc. The training images of reference classes were also collected from the Bing Image Search around the same time. Data set II uses the same testing images as in data set I. However, its training images of reference classes were collected from the Google Image Search also in July 2010. In data set III, both testing and training images were collected from the Bing Image Search but at different time (eleven months apart).7 All the testing images for re-ranking are manually labeled, while the images of reference classes, whose number is much larger, are not labeled.

References

[1] R. Datta, D. Joshi, and J.Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Computing Surveys, vol. 40, article 5, 2007.
[2] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349-1380, Dec. 2000.
[3] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra, "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval," IEEE Trans. Circuits and Systems for Video Technology, vol. 8, no. 5, pp. 644-655, Sept. 1998.
[4] X.S. Zhou and T.S. Huang, "Relevance Feedback in Image Retrieval: A Comprehensive Review," Multimedia Systems, vol. 8, pp. 536-544, 2003.
[5] 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. Pattern Analysis and Machine Intelligence, vol. 28, no. 7, pp. 1088-1099, July 2006.
[6] J. Cui, F. Wen, and X. Tang, "Real Time Google and Live Image Search Re-Ranking," Proc. 16th ACM Int'l Conf. Multimedia, 2008.
[7] J. Cui, F. Wen, and X. Tang, "Intent Search: Interactive on-Line Image Search Re-Ranking," Proc. 16th ACM Int'l Conf. Multimedia, 2008.
[8] X. Tang, K. Liu, J. Cui, F. Wen, and X. Wang, "Intent Search: Capturing User Intention for One-Click Internet Image Search," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 34, no. 7, pp. 1342-1353, July 2012.
[9] N. Rasiwasia, P.J. Moreno, and N. Vasconcelos, "Bridging the Gap: Query by Semantic Example," IEEE Trans. Multimedia, vol. 9, no. 5, pp. 923-938, Aug. 2007.
[10] C. Lampert, H. Nickisch, and S. Harmeling, "Learning to Detect Unseen Object Classes by Between-Class Attribute Transfer," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2009.
[11] Q. Yin, X. Tang, and J. Sun, "An Associate-Predict Model for Face Recognition," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2011.
[12] A. Kovashka, D. Parikh, and K. Grauman, "WhittleSearch: Image Search with Relative Attribute Feedback," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2012.

Author Profile

Dipak Pardhi is Head of Department in Computer Engineering in Godavari Foundation's Godavari College of Engineering, North Maharashtra University, Jalgaon 425001, India

Prajakta Pachpande received the B.E. degree in Computer Engineering from North Maharashtra University, Jalgaon in 2011. She is presently doing M.E. degree in Computer Engineering from North Maharashtra University, Jalgaon.