International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 2, Mar-Apr 2015

RESEARCH ARTICLE OPEN ACCESS

Review of Re-Ranking Of Images Based On Semantic Signature

Smita Patil [1], Gopal Prajapati [2]
Department of Computer Science and Engineering
RGPV Bhopal/RKDFSOE, Indore
Madhya Pradesh - India

ABSTRACT

Current commercial search engines have adopted image re-ranking in order to improve the web based image search. Firstly the images are retrieved and displayed based on textual similarity, then user is asked to select one image from that pool as query image and the remaining images are re-ranked based on their visual similarity with the query image. But here major challenge is that, the similarities f visual features of images do not properly correlate the semantic meaning of images, which interprets the users' search intention. Recently proposed systems use semantic space to match the images, which uses attributes or reference classes as basis. But it is very difficult and inefficient to learn universal visual semantic space to characterize highly diverse images on web. Here in this system, we proposed a framework, which use two phases i.e. offline and online phases. In offline phase it automatically learns different semantic spaces for different query images. The visual features of images are projected into the semantic spaces to get the semantic signatures. In online phase images are re-ranked based on the semantic signatures. This system improves accuracy and efficiency of image re-ranking.

Keywords:- Sematic Signature.

1. INTRODUCTION

In web scale image searching keywords are used as queries and surrounding textual similarities are used as basis. But this approach suffers from ambiguity, because it is very difficult for user to describe all visual features in words. 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 [2], [3] with relevance feedback [4], [5], [6] 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 [7], [8], [9], which limits users' effort to just one-click feedback, is an effective way to improve search results and its interaction is simple enough. In this approach the word image index file and visual features of images are pre-computed 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. 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. 1.

Figure. 1. All the images shown in this figure are related to palm trees. They are different in color, shape, and texture.

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 [10], [11], [12], [13]. For example, Kovashka et al. [13] 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. 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.

In this paper, a framework is proposed for web image re-ranking. It individually and automatically learns the different query keywords instead of manually defining a universal concept dictionary. In this way the semantic space related to the image to be re-ranked becomes narrow, because of the query keyword provided by users. For example, if the query keyword is “apple,” the concepts of “mountain” and “Paris” are irrelevant and should be excluded. Instead, the concepts of “computer” and “fruit” will be used as dimensions to learn the semantic space related to “apple.” The query-specific semantic spaces can more accurately model the images to be re-ranked, since they have excluded other potentially unlimited number of irrelevant concepts, which serve only as noise and deteriorate the re-ranking performance on both accuracy and computational cost.[1]

II. RELATED WORK

The key component of image re-ranking is to compute, visual similarities which reflects semantic relevance of images. Many visual features have been developed in recent year, but effective low level features are different for different query images. Therefore, Cui et al. [7], [8] classified query images into eight predefined intention categories and gave different feature weighting schemes to different types of query images. But, the web images are quite diverse and it was very difficult to cover all images in only eight categories. Also, there may be possibility that a query image may be classified into wrong category.

In order to reduce the semantic gap, query-specific semantic signature was first proposed in [17]. Kuo et al. [18] 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 [19], [20], [21]. It enlarged the query image by taking the top N images, which are visually more similar to the query image, as positive examples. But due to the semantic gap the all top N images may not be semantically constant with the query image, which in turn may decrease the performance of pseudo relevance feedback. In object retrieval, the local spatial configuration, of visual features are verified to get good set of positive example. But, in web search the relevant images may not contain same objects.

There is a lot of work done, in re-ranking of images retrieved by text-only search. Hsu et al. [14] used the Information Bottleneck (IB) principle to maximize the mutual information between search relevance and visual features. Baluja [15] proposed Visual Rank to analyze the visual link structures of images and to find the visual themes for re-ranking. Cai et al. [16] re-ranked images with attributes which were manually defined and learned from manually labeled training samples. These approaches assumed that there was one major semantic category under a query keyword. Images were re-ranked by modeling this single category with visual and textual features. Because of, the ambiguity of query keywords, there may be multiple semantic categories under one query keyword. Without query images selected by users, these approaches cannot accurately capture users’ search intention. In these approaches, all the concepts/ attributes/reference-classes are defined manually and then applied universally to all the images. They are more suitable for offline databases with lower diversity (such as animal databases [11], [22] and face databases [10]), since image classes in these databases can better share similarities. It is impractical, ineffective and infeasible for online image re-ranking because in order to model all the web images, a huge set of concepts or reference classes are required. 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 degrade the accuracy of re-ranking. However, how to automatically find such relevant concepts and use them for online web image re-ranking was not well explored in previous studies.[1]

III. METHODOLOGY

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 by using both textual and visual information. This set of keyword expansions defines the reference classes for the query keyword. The images are retrieved by search engines based on textual information, using keyword expansion again. In this way the training examples of reference class is obtained automatically. Images retrieved by the keyword expansion (“red apple”) are much less diverse than those retrieved by the original keyword (“apple”). After automatic removal of 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. Further the redundant reference classes are removed to improve the efficiency of online image re-ranking. 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

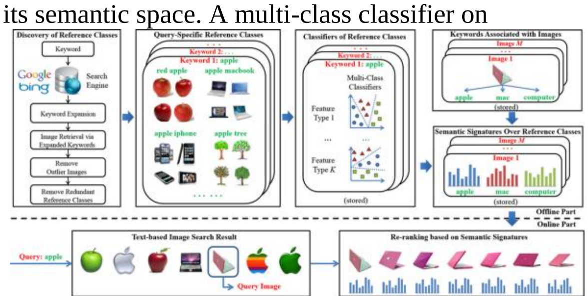

Figure 2. Our new image re-ranking framework

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. 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. 2, 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 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. Compared with the conventional image re-ranking this approach is much more efficient at the online stage, because the main online computational cost is on comparing visual features or semantic signatures and the lengths of semantic signatures are much shorter than those of low-level visual features[1].

IV. CONCLUSION

Thus here we proposed a framework. This framework uses or learns the query-specific semantic spaces in order to significantly improve the effectiveness and efficiency of online image re-ranking. The visual features of images are projected into their related semantic spaces. These semantic spaces are automatically learned through keyword expansions offline. Thus extracted semantic signatures can be 70 times shorter than the original visual features, while achieve 25-40 percent relative improvement on re-ranking. Thus the cost of online re-ranking of images is reduced to large extend.

REFERENCES

[1] Xiaogang Wang, Shi Qiu, Ke Liu, Xiaoou Tang, "Web Image Re-ranking Using Query Specific Semantic Signature", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol 36, no. 4, April 2014.

[2] 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.

[3] 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.

[4] 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.

[5] X.S. Zhou and T.S. Huang, "Relevance Feedback in Image Retrieval: A Comprehensive Review," *Multimedia Systems*, vol. 8, pp. 536-544, 2003.

[6] 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.

[7] J. Cui, F. Wen, and X. Tang, "Real Time Google and Live Image Search Re-Ranking*," Proc. 16th ACM Int'l Conf. Multimedia*, 2008.

[8] J. Cui, F. Wen, and X. Tang, "Intent Search: Interactive on-Line Image Search Re-Ranking," *Proc. 16th ACM Int'l Conf. Multimedia*, 2008.

[9] 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.

[10] 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.

[11] 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.

[12] Q. Yin, X. Tang, and J. Sun, "An Associate-Predict Model for Face Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2011.

[13] A. Kovashka, D. Parikh, and K. Grauman, "WhittleSearch: Image Search with Relative Attribute Feedback," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2012.

[14] W. Hsu, L. Kennedy, and S.F. Chang, "Video Search Reranking via Information Bottleneck Principle," *Proc. 14th Ann. ACM Int'l Conf. Multimedia*, 2006.

[15] Y. Jing and S. Baluja, "Visual Rank: Applying Page Rank to Large- Scale Image Search," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1877-1890, Nov. 2008.

[16] J. Cai, Z. Zha, W. Zhou, and Q. Tian, "Attribute-Assisted Reranking for Web Image Retrieval," *Proc. 20th ACM Int'l Conf. Multimedia*, 2012.

[17] X. Wang, K. Liu, and X. Tang, "Query-Specific Visual Semantic Spaces for Web Image Re-Ranking," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2010.

[18] Y. Kuo, W. Cheng, H. Lin, and W. Hsu, "Unsupervised Semantic Feature Discovery for Image Object Retrieval and Tag Refinement," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1079-1090, Aug. 2012.

[19] R. Yan, E. Hauptmann, and R. Jin, "Multimedia Search with Pseudo-Relevance Feedback," *Proc. Int'l Conf. Image and Video Retrieval*, 2003.

[20] R. Yan, A.G. Hauptmann, and R. Jin, "Negative Pseudo-Relevance Feedback in Content-Based Video Retrieval," *Prof. ACM Int'l Conf. Multimedia*, 2003.

[21] J. Ah-Pine, M. Bressan, S. Clinchant, G. Csurka, Y. Hoppenot, and J. Renders, "Crossing Textual and Visual Content in Different Application Scenarios," *Multimedia Tools and Applications*, vol. 42, pp. 31-56, 2009.

[22] M. Rohrbach, M. Stark, G. Szarvas, I. Gurevych, and B. Schiele, "What Helps Where and Why? Semantic Relatedness for Knowledge Transfer," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2010.