Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 5, Issue. 1, January 2016, pg.25 – 30

A Study Paper on "Required Image Retrieval by Capturing User Intention Using Image Pool Clustering"

Mr. Naresh M.Bhagat[1], Dr. M.A.Pund[2]

[1]ME (CSE), Second Year, Department of CSE, Prof. Ram Meghe Institute of Technology and Research, Badnera, Amravati
naresh.bhagat90@gmail.com
[2]Department of CSE, Prof. Ram Meghe Institute of Technology and Research, Badnera, Amravati
mapund@mitra.ac.in

Abstract— Web-scale image search engines (e.g., Google image search, Bing image search) mostly rely on surrounding text features. It is hard 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 propose a novel Internet 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 reranked based on both visual and textual content. Our key contribution is to capture 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 rerank 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 reranking. All these steps are automatic, without extra effort from the user. This is critically important for any commercial web-based image search engine, where the user interface has to be extremely simple. Besides this key contribution, a set of visual features which are both effective and efficient in Internet image search are designed. Experimental evaluation shows that our approach significantly improves the precision of top-ranked images and also the user experience.
Keywords:-"Image search, intention, image reranking, adaptive similarity, keyword expansion".

I. Introduction

Many commercial Internet scale image search engines use only keywords as queries. Users type query keywords in the hope of finding a certain type of images. The search engine returns thousands of images ranked by the keywords extracted from the surrounding text. It is well known that text-based image search suffer from the uncertainty of query keywords. The keywords provided by users tend to be small. For example, the average query length of the top 1,000 queries of image search is 1.368 words, and 97 percent of them contain only one or two words .They cannot describe the content of images precisely. The search results are noisy and consist of images with quite dissimilar semantic meaning. Therefore to solve the ambiguity, extra information has to be used to capture users' search intention. To retrieve the user intent image from image database we have to use retrieval technique. Image retrieval is the process of getting user intent image from the image database. Image clustering means grouping similar images together and comparing or matching among clusters instead of matching.

II. Literature Review

A. Improving Web-based Image Search via Content Based Clustering

Present image search engines on the web rely only on the keywords around the images and the filenames, which produces a lot of garbage in the search results. Alternatively, there exist methods for content based image retrieval that require a user to submit a query image, and return images that are similar in content. We propose a novel approach named ReSPEC (Re-ranking Sets of Pictures by Exploiting Consistency), that is a hybrid of the two methods. Our technique first retrieves the results of a keyword query from an existing image search engine, clusters the results based on extract image features, and returns the cluster that is inferred to be the most related to the search query. Furthermore, it ranks the remaining results in order of relevance.

B. *Efficient Re-Ranking of Images from the Web using Bag based Method*

An image recovery system is a computer system for browsing, searching and retrieving images from a huge database of digital images. Given a textual query in conventional text based image retrieval (TBIR), related images are to be re ranked using visual features after the preliminary text based image search. In this paper, we propose a new bag based re ranking framework for huge scale TBIR. We calculate this problem as Multiple Instance Learning and Generalized Multiple Instance learning method. To address the ambiguities on the instance labels in the positive and negative bags we propose GMI settings. Also the user log performs the procedure of individual user interface with the system which improves the performance of image retrieval.

C. Image Retrieval and Re-ranking Techniques.

There are a variety of techniques of image recovery and image re-ranking. There is a huge amount of research work focusing on the searching, retrieval and re-ranking of images in the image database. The dissimilar and spread work in this domain needs to be composed and structured for easy and quick reference. Image retrieval is a key issue of user concern. Normal way of image retrieval is the text based image retrieval technique (TBIR). TBIR-needs rich semantic textual description of web images .This technique is popular but needs very specific description of the query which is tedious and not always possible. Therefore generally the process of image search includes searching of image based on keyword typed. The process that occurs in the background is not so simple though. When query is entered in the search box for searching the image, it is forwarded to the server that is connected to the internet. The server gets the URL's of the images based on the tagging of the textual word from the internet and sends them back to the client. The search engine thus navigates through the pages and collects the images. It gives the client the top ranked image which is the one with maximum number of hits from the user and a set of images. This is the technique of text based image retrieval system. But it has certain drawbacks like images obtained are many a time duplicated, of low precision, and irrelevant.

D. Beginners to Content Based Image Retrieval.

This paper gives an overview idea of retrieving images from a large database. CBIR is used for automatic indexing and retrieval of images depending upon contents of images known as features. The features may be low level or High level. The low level features include color, texture and shape. The high level feature describes the concept of human brain. The difference between low level features extracted from images and the high level information need of the user known as semantic gap. A Single feature can represent only part of the image property. So multiple features are used to enhance the image retrieval process. This paper has used color histogram, color mean, color structure descriptor and texture for feature extraction. The feature matching procedure is based on their Euclidean distance.

III. System Analysis

In order to solve the ambiguity in text based image retrieval system. 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. Existing linguistically-related methods find either synonyms or other linguistic-related words from thesaurus, or find words frequently co-occurring with the query keywords. For example, Google image search provides the "Related Searches" feature to suggest likely keyword expansions. However, even with the same query keywords, the intention of users can be highly diverse and cannot be accurately captured by these expansions. Another way is content-based image retrieval with relevance feedback. Users label multiple positive and negative image examples. A query-specific visual similarity metric is learned from the selected examples and used to rank images. The requirement of more users' effort makes it unsuitable for web-scale commercial systems like Bing image search and Google image search in which users' feedback has to be minimized .We do believe that adding visual information to image search is important. However, the interaction has to be as simple as possible. The absolute minimum is One-Click. We propose a novel Internet image search approach. It requires the user to give only one click on a query image and images from a pool retrieved by text based search are re-ranked based on their visual and textual similarities to the query image. We believe that users will tolerate one-click interaction, which has been used by many popular text-based search engines. For example, Google requires a user to select a suggested textual query expansion by one-click to get additional results. The key problem to be solved is how to capture user intention from this one-click query image. Four steps are proposed as follows:

A. Adaptive similarity

We design a set of visual features to describe different aspects of images. How to integrate various visual features to compute the similarities between the query image and other images is an important problem. In this paper, an Adaptive Similarity is proposed, motivated by the idea that a user always has specific intention when submitting a query image. For example, if the user submits a picture with a big face in the middle, most probably he/she wants images with similar faces and using face-related features is more appropriate. In our approach, the query image is first categorized into one of the predefined adaptive weight categories, such as "portrait" and "scenery." Inside each category, a specific pretrained 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. This initial re-ranking result is not good enough and will be improved by the following steps.

B. Keyword expansion.

Query keywords input by users tend to be short and some important keywords may be missed because of users' lack of knowledge on the textual description of target images. In our approach, query keywords are expanded to capture users' search intention, inferred from the visual content of query images, which are not considered in traditional keyword expansion approaches. A word w is suggested as an expansion of the query if a cluster of images are visually similar to the query image and all contain the same word w.1 the expanded keywords better capture users' search intention since the consistency of both visual content and textual description is ensured.

C. 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 like in Google "related searches." This increases users' burden. Moreover, the suggested keywords based on text information only are not accurate to describe users' intention. Keyword expansions suggested by our approach using both visual and textual information better capture users' intention. They are automatically added into the text query and enlarge the image pool to include more relevant images. Feedback from users is not required. Our experiments show that it significantly improves the precision of top 4. Visual query expansion. One query image is not diverse enough to capture the user's intention. In Step 2, a cluster of images all containing the same expanded keywords and visually similar to the query image are found. They are selected as expanded positive examples to learn visual and textual similarity metrics, which are more robust and more specific to the query, for image re-ranking. Compared with the weight schema in Step 1, these similarity metrics reflect users' intention at a finer level since every query image has different metrics Different from relevance feedback, this visual expansion does not require users' feedback .All four of these steps are automatic with only one click in the first step without increasing users' burden. This makes it possible for Internet scale image search by both textual and visual content with a very simple user interface ranked images In system analysis the technique Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases for a recent scientific overview of the CBIR field. Content-based image retrieval is opposed to concept-based approaches. "Content-based" means that the search analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because most web-based image search engines rely purely on metadata and this produces a lot of garbage in the results. Also having humans manually enter keywords for images in a large database can be inefficient, expensive and may not capture every keyword that describes the image. Thus a system that can filter images based on their content would provide better indexing and return more accurate results. Interest in the potential of digital images has increased enormously over the last few years, fuelled at least in part by the rapid growth of imaging on the World-Wide Web. Users in many professional fields are exploiting the opportunities offered by the ability to access and manipulate remotely-stored images in all kinds' new and exciting ways. However, they are also discovering that the process of locating a desired image in a large and varied collection can be a source of considerable frustration. The problems of image retrieval are becoming widely recognized, and the search for solutions an increasingly active area for research and development.

1. Existing System:-In Existing system, one way is text-based keyword expansion, making the textual description of the query more detailed. Existing linguistically-related methods find either synonyms or other linguistic-related words from thesaurus, or find words frequently co occurring with the query keywords .For example, Google image search provides the "Related Searches" feature to suggest likely keyword expansions. However, even with the same query keywords, the intention of users can be highly diverse and cannot be accurately captured by these expansions. Search by Image is optimized to work well for content that is reasonably well described on the web. For this reason, you'll likely get more relevant results for famous landmarks or paintings than you will for more personal images like your toddler's latest finger painting.

IV. Related Work

Many Internet scale image search methods are text-based and are limited by the fact that query keywords cannot describe image content accurately. Content-based image retrieval uses visual features to evaluate image similarity. Many visual features were developed for image search in recent years. Some were global features, such as GIST and HOG. Some quantized local features, such as SIFT, into visual words, and represented images as bags-of-visual-words (BoV). In order to preserve the geometry of visual words, spatial information was encoded into the BoV model in multiple ways. For example, Zhang proposed geometry preserving visual phases which captured the local and long-range spatial layouts of visual words. One of the major challenges of content-based image retrieval is to learn the visual similarities which reflect the semantic relevance of images well. Image similarities can be learned from a large training set where the relevance of pairs of images is known. Deng learned visual similarities from a hierarchical structure defined on semantic attributes of training images. Since web images are highly diversified, defining a set of attributes with hierarchical relationships for them is challenging. In general, learning a universal visual similarity metric for generic images is still an open problem to be solved. Some visual features may be more effective for certain query images than others. In order to make the visual similarity metrics more specific to the query, relevance feedback was widely used to expand visual examples. The user was asked to select multiple relevant and irrelevant image examples from the image pool. A query-specific similarity metric was learned from the selected examples. Discriminative models were learned from the examples labeled by users using support vector machines or boosting, and classified the relevant and irrelevant images. The weights of combining different types of features were adjusted according to users' feedback. Since the number of user-labeled images is small for supervised learning methods, Huang proposed probabilistic hyper graph ranking under the semi-supervised learning framework. It utilized both labeled and unlabeled images in the learning procedure. Relevance feedback required more users' effort. For a web-scale commercial system, users' feedback has to be limited to the minimum, such as one-click feedback. In order to reduce users' burden, pseudo relevance feedback expanded the query image by taking the top N images visually most similar to the query image as positive examples. However, due to the well-known semantic gap, the top N images may not be all semantically consistent with the query image. This may reduce the performance of pseudo relevance feedback. Chum used RANSAC to verify the spatial configurations of local visual features and to purify the expanded image examples. However, it was only applicable to object retrieval. It required users to draw the image region of the object to be retrieved and assumed that relevant images contained the same object. Under the framework of pseudo relevance feedback, Ah-Pine proposed transmedia similarities which combined both textual and visual features. Krapac proposed the query-relative classifiers, which combined visual and textual information, to rerank images retrieved by an initial text-only search. However, since users were not required to select query images, the users' intention could not be accurately captured when the semantic meanings of the query keywords had large diversity. We conducted the first study that combines text and image content for image search directly on the Internet in, where simple visual features and clustering algorithms were used to demonstrate the great potential of such an approach. Following our intent image search work a visual query suggestion method is developed. Its difference and is that instead of asking the user to click on a query image for reranking, the system asks users to click on a list of keyword-image pairs generated offline using a data set from Flickr and search images on the web based on the selected keyword. The problem with this approach is that, on one hand, the data set from Flickr is too small compared with the entire Internet and thus cannot cover the unlimited possibility of Internet images and, on the other hand, the keyword-image suggestions for any input query are generated from the millions of images of the whole data set and thus are expensive to compute and may produce a large number of unrelated keyword-image pairs. Besides visual query expansion, some approaches used concept-based query expansions through mapping textual query keywords or visual query examples to high-level semantic concepts. They needed a predefined concept lexicons whose detectors were offline learned from fixed training sets. These approaches were suitable for closed databases but not for web-based image search, since the limited number of concepts cannot cover the numerous images on the Internet. The idea of learning example specific visual similarity metric was explored in previous work. However, they required training a specific visual similarity for every example in the image pool, which is assumed to be fixed. This is impractical in our application where the image pool returned by text-based search constantly changes for different query keywords. Moreover, text information, which can significantly improve visual similarity learning, was not considered in previous work.

V. Conclusion

In this system I propose a novel Internet image search approach which only requires one-click user feedback. Intention specific weight schema is proposed to combine visual features and to compute visual similarity adaptive to query images. Without additional human feedback, textual and visual expansions are integrated to capture user intention. Expanded keywords are used to extend positive example images and also enlarge the image pool to include more relevant images. This framework makes it possible for industrial scale image search by both text and visual content. The proposed new image re-ranking framework consists of multiple steps, which can be improved separately or replaced by other techniques equivalently effective. In future work, this framework can be further improved by making use of the query log data, which provides valuable co-occurrence information of keywords, for keyword expansion. One shortcoming of the current system is that sometimes duplicate images show up as similar images to the query. This can be improved by including duplicate detection in the future work. Finally, to further improve the quality of re-ranked images, we intend to combine this work with photo quality assessment work to re-rank images not only by content similarity but also by the visual quality of the images.

References

[1] Xiaoou Tang, Ke Liu, Jingyu Cui, "Intent search: Capturing User Intention For One-Click Internet Image Search," IEEE Transactions on pattern Analysis and Machine Intelligence, Vol. 34, No. 7, July 2012

[2] F. Jing, C. Wang, Y. Yao, K. Deng, L. Zhang, and W. Ma, "Igroup: Web Image Search Results Clustering," Proc. 14th Ann. ACM Int'l Conf. Multimedia, 2006.

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

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

[5] "Bing Image Search," http://www.bing.com/images, 2012.

[6] N. Ben-Haim, B. Babenko, and S. Belongie, "Improving Web-Based Image Search via Content Based Clustering," Proc. Int'l Workshop Semantic Learning Applications in Multimedia, 2006.

[7] R. Fergus, P. Perona, and A. Zisserman, "A Visual Category Filter for Google Images," Proc. European Conf. Computer Vision, 2004.

[8] G. Park, Y. Baek, and H. Lee, "Majority Based Ranking Approach in Web Image Retrieval," Proc. Second Int'l Conf. Image and Video Retrieval, 2003.

[9] Y. Jing and S. Baluja, "Pagerank for Product Image Search," Proc. Int'l Conf. World Wide Web, 2008.

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

[11] R. Datta, D. Joshi, and J.Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," ACM Computing Surveys, vol. 40, pp. 1-60, 2007.

[12] A. Torralba, K. Murphy, W. Freeman, and M. Rubin, "Context-Based Vision System for Place and Object Recognition," Proc. Int'l Conf. Computer Vision, 2003.

