Image Retrieval based on intention with Hybrid Query Model for the Web

P. Matheswaran
II-ME-CSE
M.Kumarasamy Collge of Engg.
Karur-639113
Mobile:9976699053

K. Syed Ali Fathima
LECTURER,CSE
M.Kumarasamy Collge of Engg,
Karur-639113
Mobile:9095030095

ABSTRACT

The image retrieval applications are designed to collect images based on textual query or image contents. Web-scale image search engines (e.g., Google image search, Bing image search) mostly rely on surrounding text features. 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. Internet image search approach is used to fetch images on the web environment. 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. The key contribution is to capture the users' search intention from this one-click query image. The user intent collection 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.RankBoost framework algorithm is enhanced to rank images with photographic quality. Content similarity and visual quality factors are used for the re-ranking process. Redundant image filtering process is integrated with the system. Query expansion is upgraded using query patterns and associations. The approach significantly improves the precision of top-ranked images and also the user experience.

Keywords:
Image search, intention, image reranking, adaptive similarity, keyword expansion.

1. 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 suffers from the ambiguity of query keywords. The keywords provided by users tend to be short. For example, the average query length of the top 1,000 queries of Picsearch is 1.368 words, and 97 percent of them contain only one or two words. They cannot describe the content of images accurately [1]. The search results are noisy and consist of images with quite different semantic meanings. The top ranked images from Bing image search using "apple" as query. They belong to different categories, such as "green apple," "red apple," "apple logo," and "iphone" because of the ambiguity of the word "apple." The ambiguity issue occurs for several reasons. First, the query keywords' meanings may be richer than users' expectations. For example, the meanings of the word "apple" include apple fruit, apple computer, and apple ipod. Second, the user may not have enough knowledge on the textual description of target images. For example, if users do not know "gloomy bear" as the name of a cartoon character and they have to input "bear" as query to search images of "gloomy bear." Lastly and most importantly, 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. 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. "gloomy bear" is not among the keyword expansions suggested by Google "related searches."

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.

2. Image Search Model

The method 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. In this paper, proposed 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 reranked based on their visual and textual similarities to the query image, and 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 in this paper is how to capture user intention from this one-click query image. Four steps are proposed as follows:

1. **Adaptive similarity.** The 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 rerank the text-based search result. This correspondence between the query image and its proper similarity measurement reflects the user intention. This initial reranking result is not good enough and will be improved by the following steps.

2. **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. The expanded keywords better capture users' search intention since the consistency of both visual content and textual description is ensured.

3. **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, reranking 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 ranked images.

4. **Visual query expansion.** One query image is not diverse enough to capture the user's intention. 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 reranking. Compared with the weight schema, 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. Our one-click intentional modeling has been proven successful in industrial applications and is now used in the Bing image search engine [4]. This work extends the approach to further improve the performance greatly.

3. Related Work

3.1 Image Search and Visual Expansion

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 [5], [6] 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 et al. [7] 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 [8]. Deng et al. [9] 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 et al. [10] proposed probabilistic hypergraph 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 pseudorelevance feedback. Chum et al. 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 pseudorelevance feedback, Ah-Pine et al. proposed transmedia similarities which combined both textual and visual features. Krapac et al. [2] 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.

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.

3.2 Keyword Expansion

In our approach, keyword expansion is used to expand the retrieved image pool and to expand positive examples. Keyword expansion was mainly used in document retrieval. Thesaurus-based methods expanded query keywords with their linguistically related words such as synonyms and hypernyms. Corpus-based methods, such as well-known term clustering and Latent Semantic Indexing measured the similarity of words based on their co-occurrences in documents. Words most similar to the query keywords were chosen as textual query expansion.

Some image search engines have the feature of expanded keywords suggestion. They mostly use surrounding text. Some algorithms [3] generated tag suggestions or annotations based on visual content for input images. Their goal is not to improve the performance of image reranking. Although they can be viewed as options of keyword expansions, some difficulties prevent them from being directly applied to our problem. Most of them assumed fixed keyword sets, which are hard to obtain for image reranking in the open and dynamic web environment. Some annotation methods required supervised training, which is also difficult for our problem. Different than image annotation, our method provides extra image clusters during the procedure of keyword expansions, and such image clusters can be used as visual expansions to further improve the performance of image reranking.

4. Method

4.1 Visual Feature Design

The design and adopt a set of features that are both effective in describing the visual content of images from different aspects, and efficient in their computational and storage complexity. Some of them are existing features proposed in recent years. Some new features are first proposed by us or extensions of existing features. It takes an average of 0.01 ms to compute the similarity between two features on a machine of 3.0 GHz CPU. The total space to store all features for an image is 12 KB. More advanced visual features developed in recent years or in the future can also be incorporated into our framework.

4.2 Adaptive Weight Schema

Humans can easily categorize images into high-level semantic classes, such as scene, people, or object. The method observed that images inside these categories usually agree on the relative importance of features for similarity calculations. Inspired by this observation,and assign the query images into several typical categories, and adaptively adjust feature weights within each category. Suppose an image i from query category Qq is characterized using F visual features, the adaptive similarity between image i and j is defined as sq (i, j) = ∑FM= 1∝qm Sm(i,j), where sm(i, j) is the similarity between image i and j on feature m, and ∝qm expresses the importance of feature m for measuring similarities for query images from category Qq. further constrain ∝qm ≥ 0 and

∑m ∝q m = 1.

4.3 Keyword Expansion

Once the top k images most similar to the query image are found according to the visual similarity metric, words from their textual descriptions are extracted and ranked, using the term frequency-inverse document frequency (tf-idf) method. The top m (m = 5 in our experiments) words are reserved as candidates for query expansion. Fig. 1 shows such an example. The query keyword is "palm" and the query image is the top leftmost image of "palm tree." Its top-ranked images using the adaptive weight schema are shown from left to right and from top to bottom. They include images of "palm tree", "palm treo", "palm leaves," and "palm reading." There are more images of "palm treo" than those of "palm tree," and some images of "palm tree" are ranked in low positions. Thus, the word "treo" gets the highest score calculated either by tf-idf values or tf-idf value weighted by visual distance.

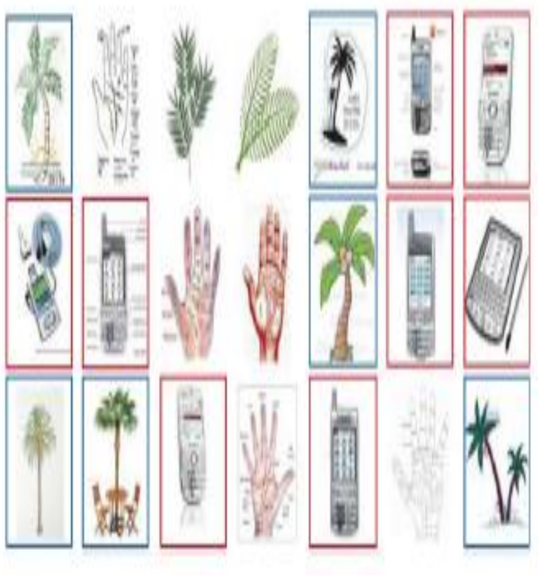

Fig. 1. Content-based image ranking result with many irrelevant images

The keyword expansion through image clustering. For each candidate word wi, all the images containing wi in the image pool are found. However, they cannot be directly used as the visual representations of wi for two reasons. First, there may be a number of noisy images irrelevant to wi. Second, even if these images are relevant to wi 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, and 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.

4.4 Visual Query Expansion

The aim at developing an image reranking method, which only requires one click on the query image and thus positive examples have to be obtained automatically. The cluster of images 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, and adopt the one-class SVM to refine the visual similarity. 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. An image to be reranked is input to the one-class SVM classifier and the output is used as the similarity (simV) to the query image. Notice the effect of this step is similar to relevance feedback. However, the key difference is that instead of asking users to add the positive samples manually, our method is fully automatic.

4.5 Image Pool Expansion

Considering efficiency, image search engines such as Bing image search only rerank 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. This method rerank the top N retrieved images by the original keyword query based on their visual similarities to the query image, and 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. This method believe that there are more relevant images in the image pool with the help of expanded query keywords.

4.6 Combining Visual and Textual Similarities

Learning a query specific textual similarity metric from the positive examples E = {e1, . . . , ej} obtained by visual query expansion and combining it with the query specific visual similarity metric can further improve the performance of image reranking. For a selected query image, a word probability model is trained from E and used to compute the textual distance distT. Let θ be the parameter of a discrete distribution of words over the dictionary. Each image i is regarded as a document di where the words are extracted from its textual descriptions.

5. Hybrid Query Model

The image search systems are designed with textual query based method. The textual query model uses the text annotation and query keyword for the image retrieval process. The content based image retrieval (CBIR) scheme uses the the query image for the retrieval process. The hybrid query model uses the textual query and query image information for the image retrieval. The keyword is collected from the user and the query image is selected from the list of initial query results. The user can select any one from the resultant images. The query is recomposed in the query expansion process. The images are reranked with reference to the query image. The final results are obtained from the image reranking process.

The image retrieval system is enhanced with various features. Image ranking with photographic quality, redundant image filtering and query expansion process. RankBoost framework algorithm is enhanced to rank images with photographic quality. Content similarity and visual quality factors are used for the re-ranking process. Redundant image filtering process is integrated with the system. Query expansion is upgraded using query patterns and associations.

The proposed system is designed with five phases. The query keyword collection is the initial phase. The query keyword is collected from the user. The image selection from the initial results is the second phase. The image that is most similar to the query text can be selected from the list of initial image results. The query expansion is the third phase. The query results are collected from the expanded query value. The query expansion is carried out using the text query and the image features. The visual and semantic image features are used in the image retrieval process. The image retrieval is the fourth phase. The photographic quality based ranking and redundant image filtering are carried out under the last phase.

6. Conclusion

Image search engines are used to search images on the Web. Description and content based image search techniques are used in the image retrieval process. The system improves the query value for the search process. Image re-ranking is initiated using the user intention image selection process. The system improves the image relevancy in search process. Minimum user effort based image search process is proposed in the system. The system supports efficient image ranking process. The system uses the query enhancement process with user intention based response model.

REFERENCES

[1] Xiaoou Tang, Ke Liu, Jingyu Cui, Fang Wen and Xiaogang Wang, "IntentSearch: Capturing User Intention for One-Click Internet Image Search" IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 34, no. 7, July 2012.

[2] L. Wu, L. Yang, N. Yu, and X. Hua, "Learning to Tag," Proc. Int'l Conf. World Wide Web, 2009.

[3] J. Krapac, M. Allan, J. Verbeek, and F. Jurie, "Improving Web Image Search Results Using Query-Relative Classifiers," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2010.

[4] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang, "Spatial-Bag-of-Features," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2010.

[5] J. Philbin, M. Isard, J. Sivic, and A. Zisserman, "Descriptor Learning for Efficient Retrieval," Proc. European Conf. Computer Vision, 2010.

[6] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large Scale Online Learning of Image Similarity through Ranking," J. Machine Learning Research, vol. 11, pp. 1109-1135, 2010.

[7] Y. Zhang, Z. Jia, and T. Chen, "Image Retrieval with Geometry- Preserving Visual Phrases," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2011.

[8] J. Deng, A.C. Berg, and L. Fei-Fei, "Hierarchical Semantic Indexing for Large Scale Image Retrieval," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2011.

[9] Y. Huang, Q. Liu, S. Zhang, and D.N. Metaxas, “Image Retrieval via Probabilistic Hypergraph Ranking,” Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, 2011.

[10] “Bing Image Search,” http://www.bing.com/images, 2012.