AN OPTIMIZATION TECHNIQUE FOR ONLINE IMAGE SEARCH RE-RANKING

[1]DIPTI BARTAKKE, [2]ARCHANA GULATI

[1]PG student University of Mumbai, [2]Asst. Prof. Department of Computer Engineering University of Mumbai
E-mail: bartakke.dipti@gmail.com, gulatiarchana81@gmail.com

Abstract- Commercial search engines such as Bing and Google have adopted Image re-ranking as an effective way to improve the results of web-based image search. Given a query keyword, a pool of images are first retrieved based on textual information. It becomes difficult for user to interpret intention only on query keywords which leads to ambiguous results which are far different from user's satisfaction. In this paper, we propose a novel Internet image search approach. The user is asked to select a query image from pool with minimum effort and images from a pool retrieved by text-based search are re-ranked based on both visual and textual content. This procedure of selecting the query image and then re-ranking requires four steps: A query image is first categorized into one of several predefined intention categories, and a specific similarity measure is used inside each category to combine image features for re-ranking based on the query image. Query keywords are expanded to capture user intention, through the visual content of the query image selected by the user and through image clustering. Image pool is enlarged to contain more relevant images. The query image is also expanded by using keyword expansion. All four of these steps are automatic with only one click in the first step without increasing user's burden. This makes it possible for Internet scale image search by both textual and visual content with a very simple user interface.

Index Terms- Adaptive similarity, Keyword Expansion, Image Pool Expansion, Visual Query Expansion.

I. INTRODUCTION ABOUT INTERNET IMAGE SEARCH

With the Internet getting available to more and more people in the last decade and with the rapidly growing number of web pages, the Internet is a vast resource of information and images. How to find just the right bit of images that user need from the Internet is a big challenge in image retrieval. Even though there are a lot of well-known search engines like Google or Yahoo, still it is sometimes not easy to find the images one is interested in. One reason for this is that many users search the internet with keywords and with the huge amount of data on the internet, almost all of these keywords will be ambiguous to a certain degree, e.g., someone searches for information about the city New York and queries a search engine with the keywords "New York" he or she will get pages belonging to different areas of interest like pages about the city itself, about the newspaper The New York Post , about the basketball team New York Knicks and so on. In this case, people will either have to look through a lot of irrelevant Pages, or he will have to refine his query to better match the topic he is interested in. Either way it will take more time to find the specific information. This problem is due to the fact that most search engines perform keywords-based matching and various ranking algorithm. Therefore, all the three mentioned concepts that are related to New York (City, newspaper and basketball) will have "New York" as a keyword and will thus be relevant in a system that uses a keyword-based retrieval approach. So using visual information to re-rank and improve text based image search results is a natural idea. In addition, these approaches cannot handle ambiguity inside a keyword query, since the assumption that images returned by querying one keyword are all from one class does not hold, and the structure of the returned image set is much more complicated. For example, using "lotus" as a query keyword, the retrieved images belong to different 3 main categories, such as "lotus flower", "lotus car", and "lotus flower tattoo". Within each main class, there can be several distinct sub classes(images that are visually similar). Also, there are images that can be labelled as noise(irrelevant images) or neglect(hard to judge relevancy).



Fig 1. Top ranked images returned from Bing image search using "lotus" as query

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. In this paper, a novel Internet image search approach is proposed. After query by keyword, user can click on one image, indicating this is the query image.

We then re-rank all the returned images according to their similarities with the query. The challenging 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:

A. Adaptive Similarity

The query image is categorized into one of the pre-defined adaptive weight categories which reflects user intention. Inside each category, a specific weight schema is used to combine visual features adaptive to this kind of images to better re-rank images. This correspondence between the query image and its proper similarity measurement reflects the user intention.

B. Keyword Expansion

The keywords provided by users tend to be short. They cannot describe the content of images accurately. 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. 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 (Fig. 2a) and they have to input "bear" as query to search images of "gloomy bear". In many cases it is hard for users to describe the visual content of target images using keywords accurately.



(a) Images of "gloomy bear"



(b) Google related searchesof query "bear"

Fig 2. (a) Images of "gloomy bear" (b) Google related searches of query "bear"

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. As shown in Fig. 2b, "gloomy bear" is not among the keyword expansions suggested by Google "related searches." 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.

C. Image Pool Expansion

The text based search retrieves the pool of images that accommodates the images with large variety of semantic meaning and the number of images related to the query image is small. So, re-ranking images in the pool is not very effective. Thus to narrow the intention and retrieve more relevant images more accurate query by keywords is needed. One way is to ask the user to click on one of the suggested keywords and to expand query results like in Google "related searches".The suggested keywords based on text information only are not accurate to describe users' intention. Instead user's intention are better captured by keyword expansion suggested by our approach using both visual and textual information. Here user's feedback is not required, they are automatically added into the text query and enlarge the image pool to include more relevant images.

D. Visual Query Expansion

The goal of visual query expansion is to obtain multiple positive example images to learn a visual similarity metric which is more robust and more specific to the query image. To capture the user's intention only one query image is not enough. These similarity metrics reflect users' intention at a finer level since every query image has different metrics. Visual query expansion also does not requires user's feedback.

II. LITERATURE REVIEW

A. Traditional Image Re-ranking Framework

Major web image search engines have adopted the strategy which is shown in Figure 3. Given a query keyword input by a user, a pool of images relevant to the query keyword are retrieved by the search engine according to a stored word-image index file. 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 pre-computed offline and stored. Visual features must be saved. The web image collection is dynamically updated. If the visual features are discarded and only the similarity scores of images are stored, whenever a new image is added into the collection and we have to compute its similarities with existing images, then the visual features need be computed again.

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.

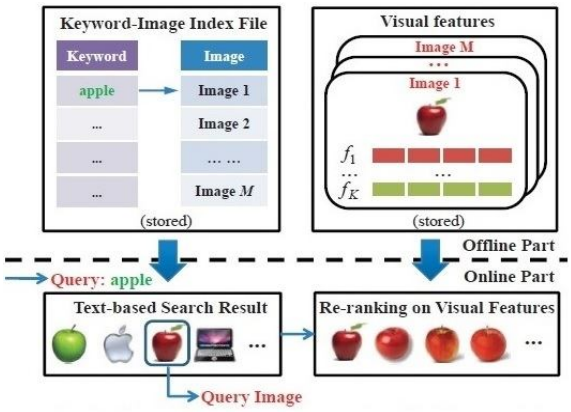

Fig 3. Traditional image re-ranking framework

B. Text-based Image Search

Many Internet scale image search methods [4], [5] are text-based and are limited by the fact that query keywords cannot describe image content accurately. In paper [4] an approach named ReSPEC (Re-ranking Sets of Pictures by Exploiting Consistency), that is a hybrid of the two methods. In paper [5] it is shown that visual consistencies in the output images can be identified, and then used to rank the images according to their closeness to the visual object category.

C. Content-based Image Search

CBIR (Content-based image retrieval) [6] uses visual features to evaluate image similarity. Drawback of CBIR is that Content-based image retrieval uses visual feature to evaluate image similarity.

D. Visual Expansion

Many visual features [7], [8], [9], [10], were developed for image search in recent years. Some were global image features, such as GIST [7] and HOG (Histogram of Oriented Gradient) [8]. Kevin proposed GIST [7] which exploit visual context, by which we mean a low-dimensional representation of the whole image.

Some local image features such as SIFT [9] David proposed a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene. In paper BoV (bags-of-visualwords) [10] the aim is to retrieve those key frames and shots of a video containing a particular object with the ease, speed and accuracy with which Google retrieves text documents containing particular words was proposed by Sivic.

Online algorithm for image similarities can be learned from a large training set where the relevance of pairs of images is known [11]. In general, learning a universal visual similarity metric for generic images is still an open problem to be solved. In order to make the visual similarity metrics more specific to the query, relevance feedback [12] was widely used to expand visual examples. The user was asked to select multiple relevant and irrelevant image examples from the image pool.

Following our intent image search work in [1] and [2] proposed by Cui, a visual query suggestion method is developed in [15] by Zha it provides a more effective query interface to formulate an intent-specific query by joint text and image suggestions. Instead of asking the user to click on a query image for re-ranking, in this system a user submits a text query, VQS (Visual Query Suggestion) first provides a list of suggestions. If the user selects one of the suggestions, the corresponding keyword will be added to complement the initial text query as the new text query, while the image collection will be formulated as the visual query.

E. 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. Corpus-based methods, such as well-known term clustering and Latent Semantic Indexing [16] proposed by Deerwester, 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.

III. PROPOSED SYSTEM

E. Approach Overview

The flow of our system is shown in Figure 4.

Step a:

First the user has to submit the query keyword q. A pool of images is retrieved by text-based search (Fig. 4a). Then the user is asked to select a query image from the image pool.

Step b:

The query image is classified as one of the predefined adaptive weight categories. Images in the pool are re-ranked (Fig. 4b) based on their visual similarities to the query image.

Step c:

The similarities are computed using the weight schema (Fig. 4c) specified by the category to combine visual features.

Step d:

In the Keyword expansion step words are extracted from the textual descriptions of top k images most similar to the query image, and tf-idf method is used to rank these words (Fig. 4d). To save computational cost, only the top m words are reserved as candidates for further processing. However, because the initial image re-ranking result is still ambiguous and noisy, the top k images may have a large diversity of semantic meanings and cannot be used as visual query expansion. The word with the highest tf-idf score computed from the top k images is not reliable to be chosen as keyword expansion either. In this

approach, reliable keyword expansions are found through further image clustering. For each candidate wordw_i, we find all the images containing w_i and group them into different clusters $\{c_{i.1}, c_{i,2}, \ldots, c_{i,ti}\}$ based on visual content. As shown in (Fig. 4d), images with the same candidate word may have a large diversity in visual content. Images assigned to the same cluster have higher semantic consistency since they have high visual similarity to one another and contain the same candidate word. Among all the clusters of different candidate words, cluster $c_{i,j}$ with the largest visual similarity to the query image is selected as visual query expansion (Fig. 4d), and its corresponding word w_i is selected to form keyword expansion $q' = q + w_i$.

Step e:

A query specific visual similarity metric and a query specific textual similarity metric are learned from both the query image and the visual query expansion.

Step f:

The image pool is enlarged through combining the original image pool retrieved by the query keywords q provided by the user and an additional image pool retrieved by the expanded keywords q' (Fig. 4f).

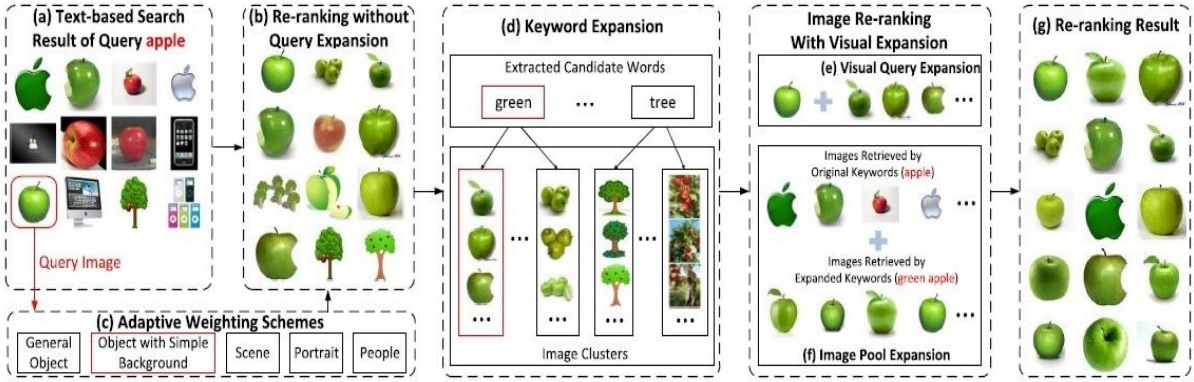

Fig 4. An example to illustrate our algorithm. The details of steps (c)-(f) are given in Sections 3.3, 3.4, 3.5, and 3.6.

Step g:

Images in the enlarged pool are re-ranked using the learned query-specific visual and textual similarity metrics (Fig. 4g). The size of the image cluster selected as visual query expansion and its similarity to the query image indicate the confidence that the expansion captures the user's search intention. If they are below certain thresholds, expansion is not used in image re-ranking.

F. Feature Design

A feature is defined to capture a certain visual property of an image, either globally for the entire image or locally for a small group of pixels. The most commonly used features include those reflecting color, texture, shape, and salient points in an image.

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 can also be incorporated into our framework.

i. Gist

The GIST descriptor focuses on the shape of scene itself, on the relationship between the outlines of the surfaces and their properties, and ignores the local objects in the scene and their relationships, and works well for scenery images.

ii. SIFT

The SIFT descriptor is invariant to translations, rotations and scaling transformations in the image domain and robust to moderate perspective transformations and illumination variations.

iii. Histogram of Gradient (HoG)

Histogram of Oriented Gradients (HOG) are feature descriptors used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image.

iv. Attention Guided Color Signature

Color signature is to describe the overall color distribution in an image. In this paper, we used an attention guided color signature, which takes the importance of pixels into consideration when calculation the color distribution. It first use an attention detector to compute a saliency map for the image, then perform k-means clustering weighed by this map.

v. Color Spatialet (CSpa)

We design a novel feature, Color Spatialet, to characterize the spatial distribution of colors. An image is divided into n x n patches. Within each patch, we calculate its main color as the largest cluster after k-Means clustering. The image is characterized by CSpa, a vector of n^2 color values. In

our experiments, we take n = 9. We account for some spatial shifting and resizing of objects in the images when calculating the distance of two CSpas A and B:
$d(A, B) = \sum_{i=1}^{n} \sum_{j=1}^{n} \min [d(A_{i,j}, B_{i\pm1,j\pm1})]$

Where $A_{i,j}$ denotes the main color of the (i, j) th block in the image.

vi. Multi-Layer Rotation Invariant EOH

Edge Orientation Histogram (EOH), which describes histogram of edge orientations, has long been used in vision applications. To calculate the distance between two MRI-EOHs, we rotate one of them to best match the other, and take this distance as the distance between the two.

vii. Facial Feature

Face existence and their appearances give clear semantic interpretations of the image. We apply face detection algorithm to each image, and obtain the number of faces, face size and position as the facial feature (Face).

G. Search By Adaptive Similarity

Given a set of images returned by text based image search engines, we try to leverage the power of visual features adaptively to re-rank the search results. Suppose we have F different features to characterize an image. The normalized similarity between image i and j on feature m is denoted as $s_i^m(j)$, which takes value in the range of [0; 1]. A vector α_i is defined for each image i to express its specific "point of view" towards different features. The larger α_{im} is, the more important the mth feature will be for image i. Without losing generality, we further constrain $\alpha \geq 0$ and $\|\alpha\|_1 = 1$, and get the adaptive similarity measurement at image i as a linear combination of its similarities on different features weighted by α_i: $s_i(.) = \sum_{m=1}^{F} \alpha_{mi} s_i^m(.)$.Our purpose is to adjust the weight αadaptively for each query image i.

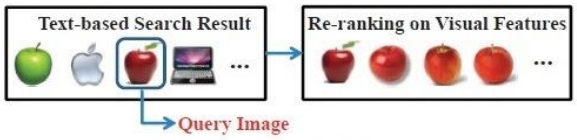

Fig 5. Re-ranking of visual features

i. Intention Categorization

Human can easily categorize images into high level semantic classes, such as scene, people, or object. Inspired by this observation, we roughly summarize general images into typical intention categories as:

1. General Object. Images containing close-ups of general objects; 2. Object with Simple Background;3. Scene. Scenery images; 4. Portrait. Images containing portrait of a single person;5. People. Images with general people inside, and are not "Portrait".
More intention categories can certainly be added in future work.

The problem of intention categorization is different from that of object categorization. On the contrary, since human beings are good at high level categorization, we can make use of this ability by designing several specific "attributes" advised by human prior to be highly related to intention categorization. Using these attributes, we are actually mapping the images into a space in which intention categorization is relatively easy. As a second step, we use C4.5 decision tree to handle the set of inhomogeneous attributes.

The attributes for intention categorization includes:
1. Face existence;2. Face number; 3. Face size; 4. Face position; 5. Directionality. Kurtosis of Edge Orientation Histogram; 6. Color Spatial Homogeneousness;7. Edge Energy; 8. Edge Spatial Distribution. With these attributes, we train a C4.5 decision tree on an image set with manually labelled intentions.

ii. Intention Specific Feature Fusion

In each intention category, we pre-train an optimal weights αto combine the visual features to achieve best re-rank performance in this specific intention based on the RankBoost Framework [18]. A ranking feature is nothing more than an ordering of the instances from most preferred to least preferred. Here our goal is instead to come up with a good ranking of the instances, so we need some feedback describing, by example, what it means for a ranking to be "good". One natural way of representing such feedback would be in the same form as that of a ranking feature. For a query image i, a real valued feedback function Φ (j, k) is defined to denote preference between image j and k. If image k should be ranked above image j, we let Φ (j, k) >0, and let Φ (j, k) = 0 otherwise. Based on Φ, a distribution over all pairs of images (j, k) is defined as D (j, k) = c Φ (j, k),where c is a constant to ensure $\sum_{j,k} D(j,k) = 1$. Thus, the ranking loss using similarity measurement $s_i(.)$ will be: $L_i = Pr_{(j,k)\sim D_i}[s_i(k) \leq s_i(j)]$.

Optimizing this loss with respect to the variable α_i in $s_i(.)$ will produce optimal weight schema for image i. To solve the problem, we use the RankBoost framework in Algorithm 1, but there are 4 steps in the procedure that varies from the original approach to accommodate our scenario: Step 1, 3, 4, and 8. we describe them below.

Algorithm 1. Feature weight learning for a certain queryCategory

1. Input: Initial weight D_i for all query images i in thecurrent intention category Q_q, similarity matrices S_m (i,.) for all query image i and feature m;
2. Initialize: Set step t=1, set $D_i^1 = D_i$ for all i;While not converged doFor each query image $i \in Q_q$ do
3. Select best feature m_t and the corresponding

similarity $s_{m_t}(i, .)$ for current re-ranking problemunder weight D_i^t;
4. Calculate ensemble weight α_{mt};
5. Adjust weight $D_i^{t+1}(j,k) \propto D_i^t(j,k)\exp\{\alpha_{mt}[s_{mt}(i,j) - s_{mt}(i,k)]\}$;
6. Normalize D_i^{t+1} to make it a distribution; 7. t++;
end for
end while
8. Output: Final optimal similarity measure for current
intension category $s^q(.,.) = \frac{\sum_t \alpha_t s_{mt}(.,.)}{\sum_t \alpha_t}$, and the
weight for feature m: $\alpha_m^q = \frac{\sum_{m_t=m} \alpha_t}{\sum_t \alpha_t}$.

Step1: Initialization. The training images are categorized into the five main classes. Images within each main class are further categorized into subclasses. Images in each subclass are visually similar. Besides, a few images are labelled as noise (irrelevant images) or neglect (hard to judge relevancy).

Step3: Select Best Feature. In step 3 of Algorithm 1, we need to select one from a set of F features that can perform best under current weight $D_i^t$for query image i. The best feature m_t is found by enumerating all F features.

Step 4: Select Ensemble Weight. It is proven in [51] that minimizing the normalization factor Z_i^tin each step is approximately equivalent to minimizing the upper bound of the rank loss. Ensemble weight is calculated by minimizing the upper bound of Z i.e.,

$\hat{Z} = \left(\frac{1-r}{2}\right)e^{\alpha} + \left(\frac{1+r}{2}\right)e^{-\alpha}$,
wherer = $\sum_{j,k} D_i(j,k)[s_i(k) - s_i(j)]$.
We know that the new $\hat{Z}$ is minimized when

$$\alpha = \frac{1}{2}\ln\left(\frac{1+r+e^{\alpha'}}{1-r+e^{-\alpha'}}\right) \qquad (1)$$

Step8: Obtain Final Weight for Feature Fusion. The final output of the new RankBoost algorithm is a linear combination of all the base rankers generated in each step.

H. Keyword Expansion

Keyword expansion is used to expand the retrieved image pool and to expand positive examples. Keyword expansion was mainly used in document retrieval. Once the top k images most similar to the query image are found, words from their textual descriptions are extracted and ranked, using the term frequency-inverse document frequency (tf-idf) [17] method. The top m (m = 5) words are reserved as candidates for query expansion. Because of the semantic diversity of the top k images, the word with the highest tf-idf score may not capture the user's search intention.

If there are many irrelevant images among the top k images, the performance of these methods is degraded. Fig. 6 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" (marked by blue rectangles), "palm treo" (marked by red rectangles), "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. We do keyword expansion through image clustering. For each candidate word $w_i$all the images containing w_i in the image pool are found. However, they cannot be directly used as the visual representations of w_i for two reasons. First, there may be a number of noisy images irrelevant to w_i. Second, even if these images are relevant to w_i semantically, they may have quite different visual content. Fig. 7 shows such an example. In order to find images with similar visual content as the query example and remove noisy images, we 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.



Fig 6. Content-based image ranking result with many irrelevant images among top ranked images. The query keyword is "palm" and the query image is the top leftmost image of "palm tree".



Fig 7. Examples of images containing the same word "palm tree" but with different visual content.

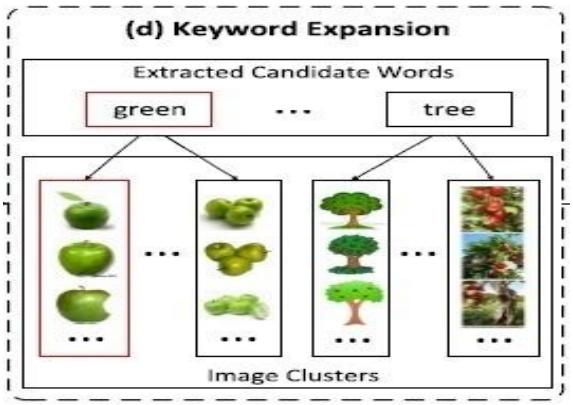

Fig 8. Keyword Expansion

Each word w_i has t_i clusters $C(w_i) = \{c_{i,1}, \ldots, c_{i,ti}\}$. The visual distance between the query image and a cluster c is calculated as the mean of the distances between the query image and the images in c. The cluster $c_{i,j}$ with the minimal distance is chosen as visual query expansion and its corresponding word w_i, combined with the original keyword query q, is chosen as keyword expansion q'. See the example in Fig. 8. If the distance between the closest cluster and the query image is larger than a threshold, it indicates that there is no suitable image cluster and word to expand the query, and thus query expansion will not be used.

I. Visual Query Expansion

So far we only have one positive image example which is the query image. The goal of visual query expansion is to obtain multiple positive example images to learn a visual similarity metric which is more robust and more specific to the query image. An example in Fig. 9 explains the motivation. The query keyword is "Paris" and the query image is an image of "Eiffel tower." The image re-ranking result based on visual similarities without visual expansion is shown in Fig. 9 and there are many irrelevant images among the top-ranked images. This is because the visual similarity metric learned from one query example image is not robust enough. By adding more positive examples to learn a more robust similarity metric, such irrelevant images can be filtered out. In a traditional way, adding additional positive examples was typically done through relevance feedback, which required more users' labelling burden. We aim at developing an image re-ranking method, which only requires one click on the query image and thus positive examples have to be obtained automatically. The cluster of images chosen in has 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. We adopt the one-class SVM to refine the visual similarity.

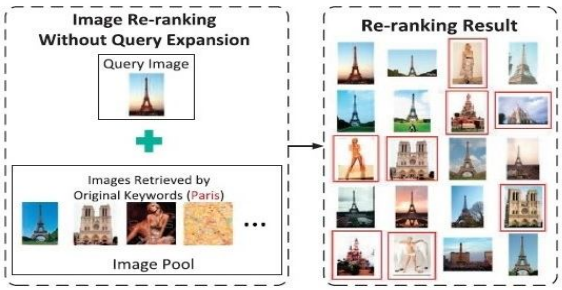

Fig 9. Image re-ranking based on visual similarity without visual query expansion

J. Image Pool Expansion

Considering efficiency, image search engines such as Bing image search only re-rank 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. In Section C, we re-rank the top N retrieved images by the original keyword query based on their visual similarities to the query image. We 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. We believe that there are more relevant images in the image pool with the help of expanded query keywords. The re-ranking result by extending image pool and positive example images is shown in Fig. 10, which is significantly improved compared with Fig. 9.

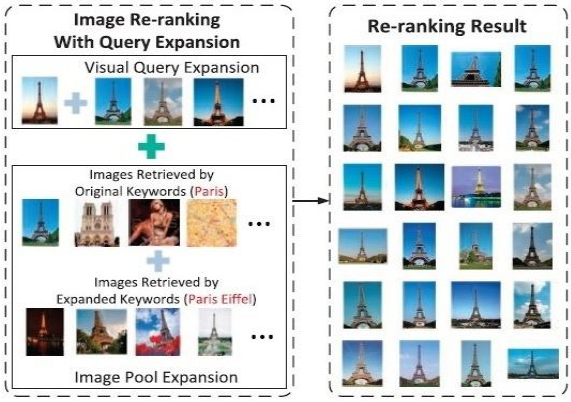

Fig 10. Image re-ranking by extending the image pool and positive example images.

K. Combining Visual and Textual Similarities

Learning a query specific textual similarity metric from the positive examples $E = \{e_1, \ldots, e_j\}$ obtained by visual query expansion and combining it with the query specific visual similarity metric introduced in Section 3.5 can further improve the performance of image re-ranking. At last, this textual distance can be combined with the visual similarity obtained in Section 3.5 to re-rank images

CONCLUSION

We have studied 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.

FUTURE WORK

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 Removal of duplicate images using histogram (Signature based) In this module, images which are been searched by the system are been compared using histogram analysis with signature based. For each and every image histogram frequency graph will be generated and based on the differences of their frequency we can come to know whether the images are same. Based on this logic we will filter the duplicate images and search will be more genuine.

REFERENCES

[1] J. Cui, F. Wen, and X. Tang, "Real Time Google and Live Image Search Re-Ranking," Proc. 6th ACM Int'l Conf. Multimedia, 2008.

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

[3] 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.

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

[5] 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.

[6] 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.

[7] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2005.

[8] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int'l J. Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[9] J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos," Proc. Int'l Conf. Computer Vision, 2003.

[10] 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.

[11] Z. Zha, L. Yang, T. Mei, M. Wang, and Z. Wang, "Visual Query Expansion," Proc. 17th ACM Int'l Conf. Multimedia, 2009.

[12] Y. Freund, R. Iyer, R.E. Schapire, and Y. Singer, "An Efficient Boosting Algorithm for Combining Features," J. Machine Learning Research, vol. 4, pp. 933-969, 2003.

★★★