Tag-Based Image Search by Social Re-ranking

Dan Lu, Xiaoxiao Liu, and Xueming Qian, *Member, IEEE*

Abstract—Social media sharing websites like Flickr allow users to annotate images with free tags, which significantly contribute to the development of the web image retrieval and organization. Tag-based image search is an important method to find images contributed by social users in such social websites. However, how to make the top ranked result relevant and, with diversity, is challenging. In this paper, we propose a social re-ranking system for tag-based image retrieval with the consideration of an image's relevance and diversity. We aim at re-ranking images according to their visual information, semantic information, and social clues. The initial results include images contributed by different social users. Usually each user contributes several images. First, we sort these images by inter-user re-ranking. Users that have higher contribution to the given query rank higher. Then we sequentially implement intra-user re-ranking on the ranked user's image set, and only the most relevant image from each user's image set is selected. These selected images compose the final retrieved results. We build an inverted index structure for the social image dataset to accelerate the searching process. Experimental results on a Flickr dataset show that our social re-ranking method is effective and efficient.

Index Terms—Image search, re-ranking, social clues, social media, tag-based image retrieval.

I. INTRODUCTION

WITH the development of social media based on Web 2.0, amounts of images and videos spring up everywhere on the Internet. This phenomenon has brought great challenges to multimedia storage, indexing and retrieval. Generally speaking, tag-based image search is more commonly used in social media than content based image retrieval [46], sketch based image retrieval [58] and context-and-content based image retrieval [51]. In recent years, the re-ranking problem in the tag-based image retrieval (TBIR) has gained researchers' wide attention.

Nonetheless, the following challenges block the path for the development of re-ranking technologies in the TBIR.

1) Tag mismatch. Social tagging requires all the users in the social network to label their uploaded images with their own keywords and share with others. Different from

ontology based image annotation, there is no predefined ontology or taxonomy in social image tagging. Every user has his own habit to tag images. Even for the same image, tags contributed by different users will be of great difference [25], [47]. Thus, the same image can be interpreted in several ways with several different tags according to the background behind the image. Thus, many seemingly irrelevant tags are introduced.

2) Query ambiguity. Users cannot precisely describe their request with single words and tag suggestion system always recommend words that are highly correlated to the existing tag set, thus add little information to a users' contribution. Besides, polysemy and synonyms are the other causes of the query ambiguity.

Thus, a fundamental problem in the re-ranking of the tag-based social image retrieval is how to reliably solve these problems. As far as the "tag mismatch" problem is concerned, tag refinement [1], [2], [20], [22], [24], [26], tag relevance ranking [17], [33], [35], [45] and image relevance ranking approach [3], [7], [15], [21], [27], [33], [34] have been dedicated to overcome this problems. As for the "query ambiguity" problem, an effective approach is to provide diverse retrieval results that cover multiple topics underlying a query. Currently, image clustering [8], [10] and duplicate removal [4]–[6], [9], [28], [29], [31] are the major approaches in settling the diversity problem. However, the essence of social images is ignored. The social images uploaded and tagged by users are user-oriented. These user-oriented images which share the same user and tagged with same query are always taken in a fixed time interval at a specific spot. It is well-known that, images taken in the same time interval and fixed spot are fairly similar. To diversify the top ranked search results, it's better to re-rank the results by removing the duplicate images from the same user.

Starting from this intuition and above analysis, we propose a social re-ranking algorithm which user information is firstly introduced into the traditional ranking method considering the semantics, social clues and visual information of images. The contributions of this paper can be described as follows.

1) We propose a tag-based image search approach with social re-ranking. We systematically fuse the visual information, social user's information and image view times to boost the diversity performance of the search result.

2) We propose the inter-user re-ranking method and intra-user re-ranking method to achieve a good trade-off between the diversity and relevance performance. These methods not only reserve the relevant images, but also effectively eliminate the similar images from the same user in the ranked results.

3) In the intra-user re-ranking process, we fuse the visual, semantic and views information into a regularization framework to learn the relevance score of every image in each

user's image set. To speed up the learning speed, we use the co-occurrence word set of the given query to estimate the semantic relevance matrix.

Comparing with the preliminary work [43], we have made some improvements as follows.

1) In order to improve the robustness of the algorithm to obtain the co-occurrence word set with respect to the given query in [43], a new self-adaptive algorithm is introduced in this paper, in which relative frequency of each tag about the given query is required and a self-adaptive parameter is decided by this relative frequency.

2) In the intra-user re-ranking process, we take the views into consideration to learn the relevance score of each image on the basis of [43]. In order to achieve this, a new iterative algorithm to obtain the relevance score is proposed.

3) Comparing with the algorithm proposed in [43], this paper is more considerate. Discussions about weight selection and image features in the regularization framework are complemented. Through this discussion, we find that our performance doesn't rely on the adjustment of parameters and feature selection. It's robust and relatively stable. Besides, in order to find an optimal number of representative images which are selected from each user's image set, many new comparison experiments and comprehensive discussions are added.

The remainder of this paper is organized as follows. In Section II, we review the related work on the re-ranking of the TBIR. The system overview is illustrated on Section III. Section IV demonstrates the offline system. The online system is depicted in Section V. Experiments on Flickr dataset are set up and shown in Section VI. Finally, conclusion and future work are given in Section VII.

II. RELATED WORK

Social image websites such as Flickr,[1] allow users to annotate their images with a set of descriptors such as tags and was often employed in image tagging, video tagging [57] and tag based image retrieval. Thus, the tag-based image search can be easily accomplished by using the tags as query terms. However, the weakly relevant tags, noisy tags and duplicated information make the search result unsatisfactory. Most of the literatures regarding the re-ranking of the TBIR focus on tag processing, image relevance ranking and diversity enhancement of the retrieval results. The following parts present the existing works related to the above three aspects respectively.

A. Tag Processing Strategy

It has been long acknowledged that tag ranking and refinement play an important role in the re-ranking of TBIR, for they lay a firm foundation on the development of re-ranking in TBIR. For example, Liu *et al.* [1] proposed a tag ranking method to rank the tags of a given image, in which probability density estimation is used to get the initial relevance scores and a random walk is proposed to refine these scores over a tag

similarity graph. Similar to [1], [2], and [26] sort the tag list by the tag relevance score which is learned by counting votes from visually similar neighbors, and the applications in TBIR also have been conducted. Based on these initial efforts, Lee and Neve [22] proposed to learn the relevance of tags by visually weighted neighbor voting, a variant of the popular baseline neighbor voting algorithm [2]. Agrawal and Chaudhary [17] proposed a relevance tag ranking algorithm, which can automatically rank tags according to their relevance with the image content. A modified probabilistic relevance estimation method is proposed by taking the size factor of objects into account and random walk based refinement is utilized. Li *et al.* [24] presented a tag fusion method for tag relevance estimation to solve the limitations of a single measurement on tag relevance. Besides, early and late fusion schemes for a neighbor voting based tag relevance estimator are conducted. Zhu *et al.* [34] proposed an adaptive teleportation random walk model on the voting graph which is constructed based on the images relationship to estimate the tag relevance. Sun *et al.* [41] proposed a tag clarity score measurement approach to evaluate the correctness of a tag in describing the visual content of its annotated images. The tag clarity score is measured by calculating the distance between the tag language model and the collection language model. Besides, many research efforts about the tag refinement emerged. Wu *et al.* [19] raised a tag completion algorithm to fill in the missing tags and correct the erroneous tags for the given image. Qian *et al.* [42] proposed a retagging approach to cover a wide range of semantics, in which both the relevance of a tag to image as well as its semantic compensations to the already determined tags are fused to determine the final tag list of the given image. Gu *et al.* proposed an image tagging approach by latent community classification and multi-kernel learning [45]. Yang *et al.* [20] proposed a tag refinement module which leverages the abundant user-generated images and the associated tags as the "social assistance" to learn the classifiers to refine noisy tags of the web images directly. In [50], Qi *et al.* proposed a collective intelligence mining method to correct the erroneous tags in the Flickr dataset.

B. Relevance Ranking Approach

To directly rank the raw photos without undergoing any intermediate tag processing, Liu *et al.* [2] utilized an optimization framework to automatically rank images based on their relevance to a given tag. Visual consistency between images and semantic information of tags are both considered. Gao *et al.* [7] proposed a hypergraph learning approach, which aims to estimate the relevance of images. They investigate the bag-of-words and bag-of-visual words of images, which are extracted from both the visual and textual information of image. Chen *et al.* [21] proposed a Support Vector Machine classifier per query to learn relevance scores of its associated photos. Wu *et al.* [15] proposed a two-step similarity ranking scheme that aims to preserve both visual and semantic resemblance in the similarity ranking. In order to achieve this, a self-tune manifold ranking solution that focuses on the visual-based similarity ranking and a semantic-oriented similarity re-ranking method are in-

---

cluded. Hu *et al.* [27] proposed an image ranking method which represent images by sets of regions and apply these representations to the multiple-instance learning based on the max margin framework. Yu *et al.* [35] proposed a learning based ranking model, in which both the click and visual feature are adopted simultaneously in the learning process. Specially, Haruechaiyasak *et al.* [33] proposed a content-based image retrieval method to improve the search results returned by TBIR. In order to give users a better visual enjoyment, Chen *et al.* [18] focused their attention on how to boost the quality of the retrieval images. They proposed a relevance-quality ranking method with the consideration of image relevance and quality value.

C. Diversity Enhancement

The relevance based image retrieval approaches can boost the relevance performance, however the diversity performance of searching are often ignored. Many researchers dedicated their extensive efforts to solve this problem. In [8], Cai *et al.* proposed a hierarchical clustering method to cluster the search results into different semantic clusters by using visual, textual and link analysis. Similarly, in [10], Leuken *et al.* studied three visually diverse ranking methods to re-rank the image search results based on the visual characteristics of these images. Different from clustering, Song *et al.* [11] proposed a re-ranking method to meet users' ambiguous needs by analyzing the topic richness. Yang and Wang *et al.* [4], [5] proposed a diverse relevance ranking algorithm to maximize average diverse precision in the optimization framework by mining the semantic similarities of social images based on their visual features and tags. Sun *et al.* [28] proposed a social image ranking scheme to retrieve the images which meet the relevance, typicality and diversity criteria by exploring both semantic and visual information of images on the basis of [5]. Ksibi *et al.* [31] proposed to assign a dynamic trade-off between the relevance and diversity performance according to the ambiguity level of the given query. Based on [31], Ksibi *et al.* [6] proposed a query expansion approach to select the most representative concept weight by aggregating the weights of concepts from different views, using a dynamic threshold. Wang *et al.* [29] proposed a duplicate detection algorithm to represent images with hash code, so that large image database with similar hash codes can be grouped quickly. Qian *et al.* [48] proposed an approach for diversifying the landmark summarization from diverse viewpoints based on the relative viewpoint of each image. The relative viewpoint of each image is represented with a 4-dimensional viewpoint vector.

However, most of the existing approaches highly rely on the visual and semantic information, and thus ignore the social clues such as user and view information. User information is always exploited to do the target advertisement [49], travel recommendation [47], personalized service for recommendation [55], [56], [59] and user interaction based image re-ranking [53]. However, user information is seldom used in retrieval work. In this paper, we propose a social re-ranking method which fuses the user information into the traditional TBIR framework. We first get the initial results by keyword matching process. Then the inter-user and intra-user re-ranking are introduced to
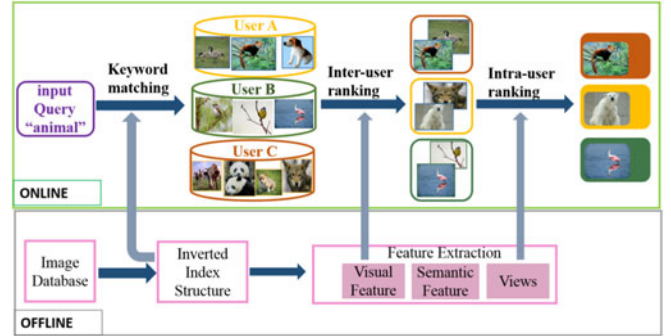

Fig. 1. System framework of tag-based image retrieval with social re-ranking.

re-rank the initial results. Inter-user re-ranking algorithm is applied to rank users according to their contribution to the given query. After the inter-user re-ranking, we further introduce intra-user re-ranking to sequentially select the most relevant image from each image dataset of the ranked users. That's to say, the final retrieved images all have different user. The most relevant image uploaded by the highest contribution user is the first in the retrieved results. Experimental results demonstrate that the proposed scheme is able to boost the diversity and relevance performance simultaneously.

Our social re-ranking system includes two main sections: online and offline as shown in Fig. 1. The offline section contains two parts: 1) Inverted index structure construction for image dataset. An inverted index structure is built to accelerate the retrieval speed. 2) Feature extraction. In this paper, we extract the visual feature, semantic feature and views for the images dataset. Semantic feature refers to the co-occurrence word set of query tags and the tags of the images.

III. SYSTEM OVERVIEW

Our online parts consist of the following three steps: 1) Keyword matching. For an input query, our system will return the initial retrieval results by keyword matching. And the following two online steps are all conducted to re-rank the initial results. 2) Inter-user re-ranking. The inter-user re-ranking is applied to rank the corresponding users with the consideration of their contributions to the given query. 3) Intra-user re-ranking. A regularization framework is proposed to determine the relevance level of each image by fusing the visual, semantic and views information into a unified system. Then we sequentially select the most relevant image in each ranked user's image set. These selected images constitute our re-ranking results. Hereinafter the details are displayed.

IV. THE OFFLINE SYSTEM

There are two main processes in the offline system: the construction of inverted index structure and the feature extraction of the image database. The details are as follows.

A. Inverted Index Structure Construction

To realize fast retrieval, an inverted index structure for the collected images is built. In our experiment, our image dataset

is composed of 6 600 034 images uploaded by 7249 users which are crawled from the public API of Flickr [1]. Each user has uploaded several images. The organization form of original images is based on users. And the inverted index structure is based on tags and each tag corresponds to the images uploaded by different users. Let G denote the total number of tags in our image dataset and the corresponding tag set is denoted by $\Gamma = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_G\}$. Γ_i denotes the ith tag that users have used to annotate their shared photos in social community. The inverted structure of the image dataset is described as $ID = \{ID_1, ID_2, \ldots, ID_G\}$. ID_i is the image collection of tag Γ_i. That is to say, all images in ID_i have been tagged with Γ_i.

B. Feature Extraction

In this paper we use the visual features, views and semantic features to represent the images in our image dataset.

1) Visual Feature: Color feature is one of the most widely used visual features in image retrieval, for its invariance with respect to image scaling, rotation, translation. In this paper, an image is divided into four equal sized blocks and a centralized image with equal-size. For each block, a 9-D color moment is computed, thus the dimension of color comment for each image is 45. The 9-D color moment of an image segment is utilized, which contains values of mean, standard deviation and skewness of each channel in HSV color space.

Texture feature describes the structure arrangement of surfaces and their relationship to the environment, such as fruit skin, clouds, trees, and fabric. The texture feature in our method is described by hierarchical wavelet packet descriptor (HWVP) [12], [13]. A 170- D HWVP descriptor is utilized by setting the decomposition level to be 3 and the wavelet packet basis to be DB2.

In this paper, a 215-dimensional visual vector is utilized, including a 45-dimensional color moment feature, and a 170-dimensional texture feature vector. In our experiments, we also give a comprehensive discussion on utilizing the low-level features and deep learning feature, i.e. feature learned by AlexNet [52].

A similarity matrix W whose element w_{ij} is introduced to measure the visual distance between the two images i and j, with their visual features v_i and v_j. Here, w_{ij} can be directly calculated using Gaussian kernel function with a radius parameter σ as follows:

$$w_{ij} = \exp\left(-\frac{||v_i - v_j||^2}{2\ \sigma^2}\right) \qquad (1)$$

where $|| \cdot ||^2$ stands for the l_2-norm of the vector. Furthermore, σ represents the radius parameter which is set to be the mean value of all pairwise Euclidean distance between images.

2) Views Feature: The views of an image in social media community is an important feature which indicates the click count of this image. The number of click count has been utilized to improve the relevance performance of the image retrieval results [35], [38], [39]. Besides, clicks have also been used to estimate the documents relevance [36], [37]. For images in

Fig. 2.   An exemplary image from Flickr and its associated information.

Flickr, the number of click count on Flickr has been regarded as an indicator of image popularity [32]. For each image in Flickr, we can discover the associated ⟨views⟩ information of images from Fig. 2. The number demonstrates that this image has been clicked 989 times after sharing. To a given query, the higher views, the more popular and relevant the image will be. Let view_i represents the view times of the image i, its normalized form vt_i can be described as follows:

$$vt_i = \frac{\text{view}_i - \text{view}_{\min}}{\text{view}_{\max} - \text{view}_{\min}} \qquad (2)$$

where $\text{view}_{\max}$ and $\text{view}_{\min}$ are the maximum and minimum views of the images which share the same user with image i in our Flickr dataset.

3) Semantic Relevance Measurement: Co-occurrence is a linguistics term that can either mean concurrence/coincidence. In a more specific sense, co-occurrence means two terms which often appeared in the text corpus in a certain order. It can also be interpreted as an indicator of interdependency, semantic proximity or an idiomatic expression and often be used in the study of image tagging [45]. Suppose the co-occurrence word set about query q, is $S(q) = \{s_1, s_2, \ldots, s_l\}$, where l is the number of co-occurrence word with respect to the query q. Thus, we can obtain the co-occurrence word set for our tag set $S(q) = \{s_1, s_2, \ldots, s_l\}$ as follows.

1) For the tag q, we rank each tag in our dataset in a descending order of their relative frequency with respect to query q. And this ranked tag set we denote it as tag set T (q is not included). Each tag's relative frequency of query q is the number of images which tagged with tag q and itself in our image dataset.

2) In order to remove the noisy tags and eliminate the influence of the seldom-used tags, each tag's relative frequency of q is taken into account. We choose the top v tags in the tag set T as the co-occurrence word set of query q by the following rule: the difference between the relative frequency of the vth tag and the relative frequency of the $(v+1)$ th tag is maximum in tag set T.

After the co-occurrence words selection, we obtain a co-occurrence tag list for each query q, for example: sky, sun and cloud; coast, sand, ocean, and sea; band and concert; airplane, airport and aircraft; and so on.

However, each element in $S(q)$ has different importance in boosting the relevance performance of retrieval results for the query q. For example, cloud and blue are the two co-occurrence

words of sky. While cloud is more important than blue with respect to the query sky, since it plays a bigger role in identifying the sky. Therefore, we assign each co-occurrence word s_i a weight M_i based on the co-occurrence similarity, which are defined based on Google distance [11] as follows:

$$M_i = \exp\left\{-\frac{\max\{\log R(q), \log R(s_i)\} - \log R(q, s_i)}{\log(N) - \min\{\log R(q), \log R(s_i)\}}\right\}$$
(3)

where N is the number of images in our image dataset, $R(q, s_i)$ is the number of images which tagged with q and s_i in our image dataset.

As shown in Fig. 2, the tags associated with an image are arranged in a random order, which limits the effectiveness of TBIR. So, we need to measure the semantic relevance between the query and image. Thus, a semantic relevance matrix C is put forward to measure the semantic relevance between query tagged image and the query. Based on the above statement, we define the average co-occurrence similarity between the query q and the tag set of image i as C_i which is calculated as follows:

$$C_i = \frac{1}{\sum_{m=1}^{l} \text{sign}(s_m)} \sum_{m=1}^{l} \text{sign}(s_m) * M_m$$
(4)

where $\text{sign}(s_m)$ denotes whether the image i contains tag s_m or not, i.e.

$$\text{sign}(s_m) = \begin{cases} 1, & \text{if image } i \text{ is tagged with } s_m \\ 0, & \text{otherwise.} \end{cases}$$
(5)

V. ONLINE SYSTEM

Our online system carries out the following three steps to obtain the ranked images for the query tag q: 1) keyword matching, 2) inter-user re-ranking, and 3) intra-user re-ranking. The details of these three main parts in the online system will be described as follows.

A. Keyword Matching

For the query $q \in \Gamma$, from the inverted file index $ID = \{ID_1, ID_2, \ldots, ID_G\}$, we can obtain the corresponding images that all tagged with query q, which is denoted by X. It can be further described by taking the social user's information into account, as follows:

$$X = \{X(u_1), \ldots, X(u_Z)\} = \{X_1, \ldots, X_Z\}$$
$$= \{\{x_{11}, x_{12}, \ldots, x_{1N_1}\}, \ldots, \{x_{Z1}, x_{Z2}, \ldots, x_{ZN_Z}\}\}$$
(6)

where $U = \{u_1, u_2, \ldots, u_Z\}$ is the user set in the image dataset X, Z is the total number of users in X; X_i or $X(u_i)$ represents the images uploaded by the user u_i; x_{ij} is the jth image in image dataset X_i; N_i denotes the number of images in X_i.

B. Inter-User Re-ranking

After the process of keyword matching, each user is ranked by inter-user ranking. This ranking is based on the user's contribution to the given query q. Larger contribution users can show viewers more professional images about this query. And this contribution is measured upon the number of its images in X which is also tagged with words in $S(q)$.

For a user u_h, $h \in (1, \ldots, Z)$, we calculate its contribution to the query (denoted by E_h) as follows:

$$E_h = \sum_{j=1}^{k} \text{sign}(x_{hj})$$
(7)

where k is the total number of images in $X_h, \ldots$ $\text{sign}(x_{hj}) = 1$ means that the image x_{hj} is tagged with word in $S(q)$, while $\text{sign}(x_{hj}) = 0$ means the image is not.

$$\text{sign}(x_{hj}) = \begin{cases} 1, & t_{hj} \cap S(q) \neq \emptyset \\ 0, & \text{otherwise} \end{cases}$$
(8)

where t_{hj} is user annotated tag set for the image x_{hj}.

Then, we rank E_h, $h \in (1, 2, \ldots, Z)$ in a descending order. The larger of E_h, the higher corresponding user ranks, which means the user has larger contribution to the query.

C. Intra-User Re-ranking

After inter-user re-ranking, the largest contribution user is ranked highest. Then we implement intra-user re-ranking to select the image which has the highest relevant score among each user's image set. We take the image set $X_h, h \in (1, 2, \ldots, Z)$ as an example to demonstrate our intra-user re-ranking process.

For the k images in $X_h = \{x_{h1}, x_{h2}, \ldots, x_{hk}\}$, we denote their relevance scores to the query q as $\boldsymbol{r} = [r_1, r_2, \ldots, r_k]$, where r_i is the relevance score of image i to the query q. In order to obtain the r, we propose a regularization framework which fuses the visual, semantic and views information into our intra-user re-ranking approach. It is an improved version over [16]. Our regularization framework is defined as follows:

$$Q(\boldsymbol{r}) = \sum_{i,j=1}^{k} w_{ij}\left\{r_i/\sqrt{D_{ii}} - r_j/\sqrt{D_{jj}}\right\}^2$$
$$+ \alpha \sum_{i=1}^{k} (r_i - C_i)^2 + \beta \sum_{i=1}^{k} (r_i - vt_i)^2$$
(9)

where $Q(\boldsymbol{r})$ is the cost function; r_i is the relevance score of image i, $D_{ii} = \sum_{j=1}^{k} w_{ij}$, w_{ij} is the visual distance of image i and j, which is obtained in (1). C_i is the semantic relevance score of image i, which is obtained by (4). vt_i is the normalization views of image i, which is obtained by (2). This cost function consists of three items. The first term in the right-hand side means that the relevance score of visually similar images should be close, the second term and the third term are fitting constraints, which means that the relevance score is biased with preference to the semantic relevance measurements and views measurements. The trade-off between these three competing constraints is captured by two positive parameters α and β.

We aim to solve the optimization problem to get the relevance score of each image in X_h as follows:

$$\boldsymbol{r}^* = \text{argmin}(Q(\boldsymbol{r})).$$
(10)

To solve $\boldsymbol{r}$, (9) can be rewritten as the matrix form

$$Q(\boldsymbol{r}) = \boldsymbol{r}^T \left(\mathbf{I} - \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{W} \boldsymbol{D}^{-\frac{1}{2}}\right) \boldsymbol{r} + \alpha||\boldsymbol{r} - \boldsymbol{C}||^2 + \beta||\boldsymbol{r} - VT||^2 \tag{11}$$

where $\boldsymbol{D} = \mathrm{Diag}(D_{11}, D_{22}, \ldots, D_{kk})$, $\boldsymbol{C} = [C_1, C_2, \ldots, C_k]$, and $VT = [vt_1, vt_2, \ldots, vt_k]$, $\mathbf{I}$ is a unit matrix with dimension $k \times k$.

Alternatively, we can use iterative optimization algorithm to solve this problem, which avoids the intensive computation brought by the direct matrix inversion in (11). The detailed steps are shown in next.

The optimization process of regularization framework (9)

INPUT:
$\boldsymbol{C} = [\boldsymbol{C_1}, \boldsymbol{C_2}, \ldots, \boldsymbol{C_k}]$;/*semantic relevance score*/
$\boldsymbol{VT} = [\boldsymbol{vt_1}, \boldsymbol{vt_2}, \ldots, \boldsymbol{vt_k}]$;/* normalization views */
$\boldsymbol{\alpha}, \boldsymbol{\beta}$; /* two positive parameters */
$\boldsymbol{w_{ij}} \epsilon \boldsymbol{W}$ /;*the visual distance of image i and j*/
$\boldsymbol{D_{ii}}$;
$r(0) = [1/2, 1/2, \ldots, 1/2]$; /* initial value of r*/
$r(\mathrm{t})$;/*the tth iteration of r*/
OUTPUT:
$\boldsymbol{r^*} = [\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots, \boldsymbol{r_k}]$;/*the final relevance score*/
BEGIN
$t = 0$; /*iteration number*/
While($|r(t) - r(t-1)| > 0.01$ && $t < 10000$)
$\boldsymbol{r(t+1)} = \frac{1}{1+\boldsymbol{\alpha}+\boldsymbol{\beta}} \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{W} \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{r(t)} + \frac{\boldsymbol{\alpha}*\boldsymbol{C}+\boldsymbol{\beta}*\boldsymbol{VT}}{1+\boldsymbol{\alpha}+\boldsymbol{\beta}}$;
$t = t+1$;
end
return r

From above, the optimization relevance score $\boldsymbol{r^*}$ in $X_h, h \in (1, 2, \ldots, Z)$ can be achieved. Then we select the image of the highest one among $X_h, h \in (1, 2, \ldots, Z)$ as the representative image of the user u_h, which denoted by x_{fh}. Finally, we re-rank the image set $\{x_{f1}, x_{f2}, \ldots, x_{fZ}\}$ by the order of their users obtained in the inter-user re-ranking process, and get our final ranked image list.

VI. EXPERIMENTS

In order to demonstrate the effectiveness of the social re-ranking (denoted by SR) approach, we conduct experiments on our crawled Flickr images by utilizing the following 20 tags as queries: airplane, beach, Beijing, bird, blue, buildings, Christmas, cityscape, forest, reflection, garden, girl, honeybee, insect, lotus, ocean, orange, sea, sky, and zebra. We systematically make comparisons for the following seven TBIR approaches:

1) VR: view-based re-ranking, a measure that rank the initial results by views in a descending order;
2) VUR: view and user based re-ranking. This approach is based on VR, and the final re-ranked results are obtained by removing the images which share the same user. That is to say, we only keep the image with the largest views for a user in the top ranked results;

TABLE I
ILLUSTRATION OF IMAGE DATASET

		With tag	With views	With tags+ views
User	Number	7090	7241	7069
	Percentage	97.81%	99.48%	97.52%
Image	Number	5 325 265	6 593 090	5 318 503
	Percentage	80.69%	99.90%	85.58%

3) RR: relevance-based re-ranking [3], an optimization framework is applied to automatically re-rank images based on visual and semantic information;
4) CRR: co-occurrence relevance re-ranking. In this algorithm we replace the semantic relevance score in [3] with the semantic relevance score proposed in our paper. The semantic relevance score in [3] takes all the tags of images into consideration. Our proposed approach only considers the co-occurrence tags;
5) DRR: diverse relevance re-ranking [5], which optimizes an ADP measure with the consideration of the semantic and visual information of images; and
6) SR: social re-ranking. Our proposed approach dedicates to promote the relevance and diversity performance of our results. User information is utilized to boost the diversity performance. A regularization framework which fuses the semantic, visual and views information is introduced to improve the relevance performance.

A. Dataset

In order to evaluate the performance of our method, we randomly crawled more than 6 million images together with their associated information from the image sharing websites Flickr.com through its public API. The initial data includes 6 600 034 images uploaded by 7249 users and their related files recoding the information of tags and views information. We have made a statistic about all the images and users in Table I. We remove the images that have no views and no tags. Finally there are 5 318 503 images and 7069 users left.

B. Performance Evaluation

The performance evaluation of our method is voted by five volunteers who are invited to assign the relevance scores and diversity scores for the top n images of each query under different methods. The averaged relevance score is used to measure the correlation between the query and the retrieval results. And the averaged diversity score shows the diversity level of the retrieval results.

Five volunteers are asked to give the relevance score of each image among the top n results into the following four categories: 3-perfect, 2-good, 1-so so, 0-irrelevant, according to their judgment for the compared re-ranking approaches. Then, the relevance score of the image i is obtained by averaging the assigned relevance values. Let rel_i denote the relevance value of image i. The five volunteers are also asked to give the diversity score of the top n results into four categories: 3-excellent, 2-good, 1-so
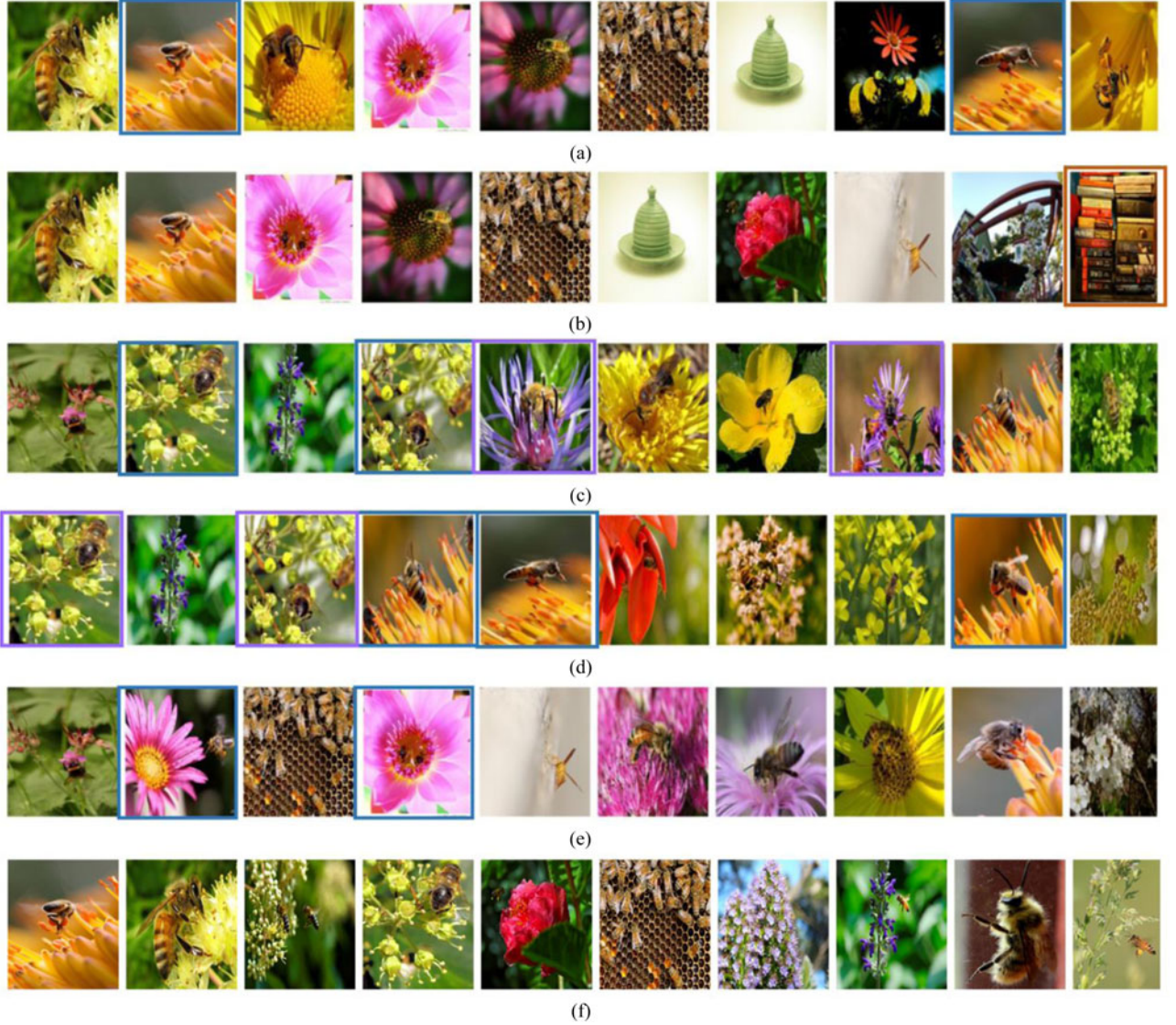
Fig. 3. Top 10 ranking results of different methods for query *honeybee*. (a) Search results using VR. (b) Searching results using VUR. (c) Searching results using RR. (d) Searching results using CRR. (e) Searching results using DRR. (f) Searching results using SR.

so, 0-similiar, according to their judgments for the compared six re-ranking approaches. Similarly, the diversity score (denoted by div@n) is obtained by averaging the assigned diversity values. The larger of the div@n, the better diversity performance is achieved.

1) Criteria of Performance Evaluation: We use the NDCG [45] and average precision under depth n (denoted as *AP@n*) as the relevance performance evaluation measure which are expressed as follows:

$$\text{NDCG@}n = \frac{1}{W} \sum_{i=1}^{n} \frac{2^{\text{lev}(i)} - 1}{\log(1+i)} \tag{12}$$

$$\text{AP@}n = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{i} \frac{\text{rel}_j}{i} \right) \tag{13}$$

where W is a normalization constant that is chosen so that the optimal ranking's NDCG score is 1.

Moreover, we can get the average diverse precision under depth n (denoted as ADP@n) as follows:

$$\text{ADP@}n = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{i} \frac{\text{rel}_j}{i} \right) *\text{norm\_div@}n \tag{14}$$

where $\text{norm\_div@}n$ is the normalized diversity value under depth n, which is represented as follows:

$$\text{norm\_div@}n = \frac{\text{div@}n}{3}. \tag{15}$$

2) Exemplar Search Results: The top 10 results of exemplar queries: honeybee, and zebra on Flickr database under six different ranking algorithms are shown in Figs. 3 and 4 respectively. The images marked with the red borders are irrelevant with the
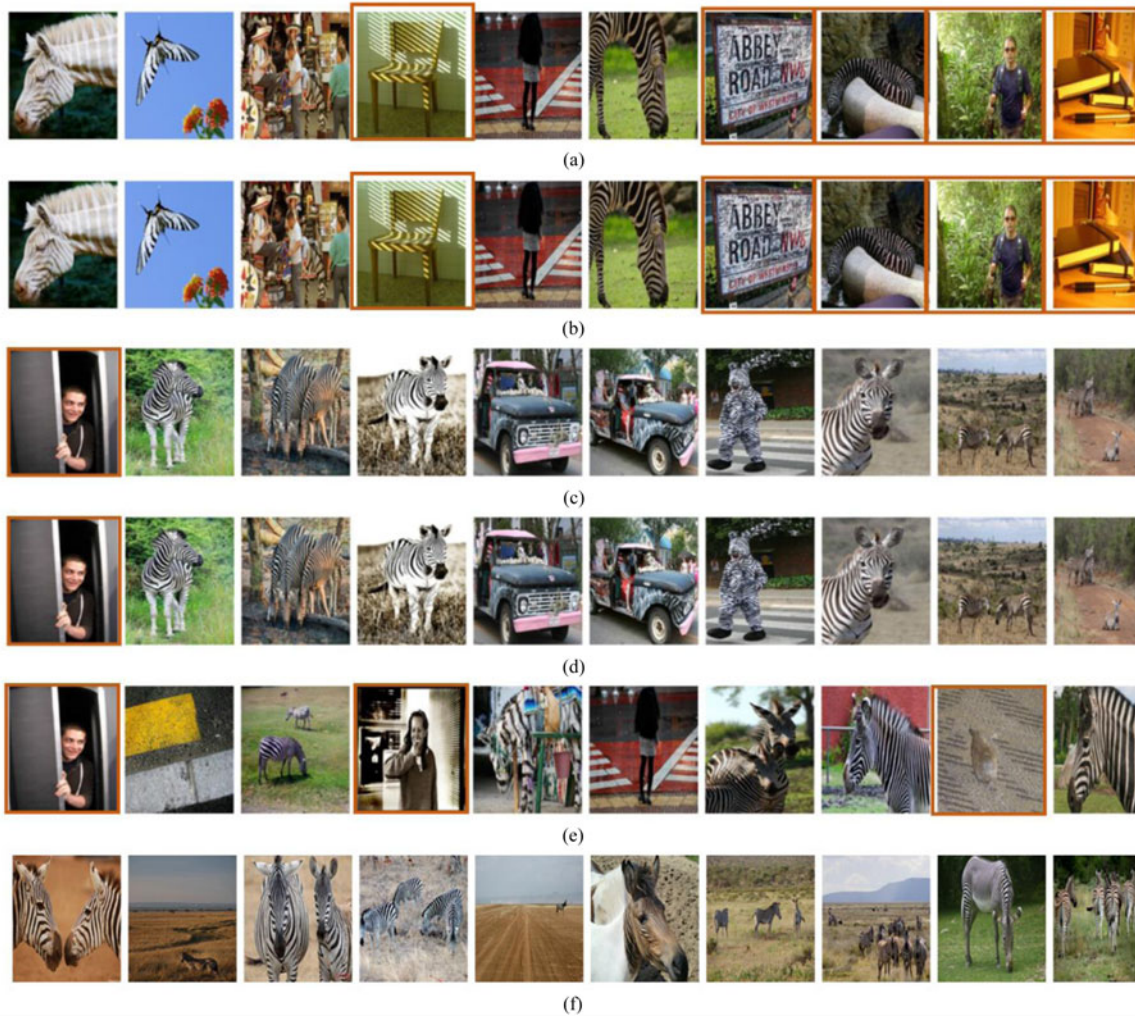
Fig. 4. Top 10 ranking results of different methods for query *zebra*. (a) Searching results using VR. (b) Searching results using VUR. (c) Searching results using RR. (d) Searching results using CRR. (e) Searching results using DRR. (f) Searching results using SR.

query. Besides, we mark the similar images from the same user with the same color.

We find that the same user's images about a same topic are always taken in the same spot at a specific time. So these images have a large probability to share the same visual appearance, tags and similar views. Therefore, the top ranked images determined by VR, RR, and CRR, are all suffered from the lack of diversity. We find that many of the relevant images obtained through them are from the same user. For example, in the search results of VR as shown in Fig. 3(a), the second and the ninth one are from the same user. For results of RR as shown in Fig. 3(c), the second and the forth, and the fifth and the eighth are from the same user. For results of CRR as shown in Fig. 3(d), the first and the third, and the fourth, the fifth and the ninth are from the same user. However, SR moves these similar images successfully. By comparing the experimental results, we find that the results of VUR and SR which introduce the social user factors and select only one representative image from same user's image set are more diverse. Additionally, from Fig. 4(a), we can also find that high views images are not all relevant

with the query q, beautiful images and images of hot topics all have a high views. The DRR introduces the semantic similarity restriction to enhance the diversity performance which brings about the promotion of the diversity performance and declines of their relevance performance, just as the result of DRR on query zebra have shown. From Fig. 4(a)–(e), we find that there are some irrelevant images in the top ranked results, just as the images with the red border shown. From the examples as shown in Figs. 3 and 4, we can acknowledge that our method takes the above deficiencies into consideration and makes a better trade-off between the diversity and relevance performance.

3) Performance Analysis: To make fair comparisons for the methods VR, VUR, RR, CRR, DRR and SR, the parameters α is all set to be 10, and β is all set to be 1. The discussions on α and β are illustrated in Section VI-D.

Let MAP@n and MADP@n denote the mean values of AP@n and ADP@n for all the 20 query tags. The NDCG@n, MAP@n and MADP@n with n = 1,5,10,15, and 20 are shown in Figs. 5, 6 and 7 respectively. For example, the MAP@20 of VR, VUR, RR, CRR, DRR, and SR are 2.52, 2.50, 2.71, 2.77, 2.64, and
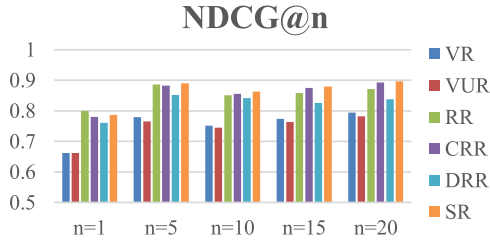
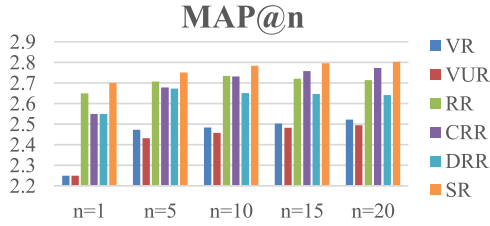Fig. 5. NDCG of all six ranking methods under different depths.

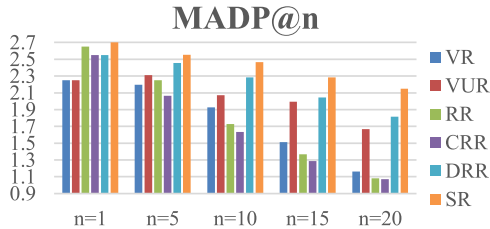Fig. 6. MAP of all six ranking methods under different depths.

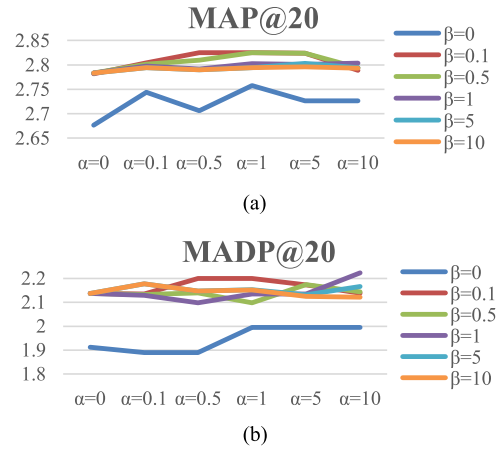Fig. 7. MADP of all six ranking methods under different depths.

(a)

(b)

Fig. 8. Impact of parameters to retrieval performance. (a) MAP@20 and (b) MADP@20 of SR under $\beta = \{0,0.1,0.5,1,5,10\}$ with fixed $\alpha = \{0,0.1,0.5,1,5,10\}$.

on the performance of our proposed image re-ranking method is discussed. Fig. 8 demonstrates the comprehensive discussions for SR under $\alpha = [0,0.1,0.5,1,5,10]$ and $\beta = [0,0.1,0.5,1,5,10]$.

As can be seen, the MADP@20 of SR with fixed $\beta = 0$ (under the case that $\alpha \neq 0$ and $\beta = 0$) is the biggest when $\alpha = 1, 5, 10$; the MADP@20 of SR under fixed $\alpha = 0$ (under the case that $\alpha = 0$ and $\beta \neq 0$) is the same under each various β. Hence, the parameter α in RR, CRR, DRR is set to be 10. The statistics in Fig. 8 show that the SR achieves the highest performance under $\alpha = 10, \beta = 1$.

From Fig. 8(a) and (b) we find that our approach under the case that $\alpha = 0$ and $\beta = 0$ is with lowest performances. It means that only from the visual information the image ranking performances are not satisfactory. When utilizing the semantic information but without the view information (under the case that $\alpha \neq 0$ and $\beta = 0$) in intra-user re-ranking, some improvements are achieved. It reflects that user's views can gain a larger performance on the help of inter-user re-ranking. When utilizing the semantic information and the view information (under the case that $\alpha \neq 0$ and $\beta \neq 0$) in intra-user re-ranking, best improvement achieved. This is likely caused by the following two aspects: 1) The user marked view information can be viewed as high level semantic information which is important in image retrieval. 2) The semantic information extracted from the user annotated tags is not robust, which may be disturbed by noise tags or user's own vocabularies [25]. By combining both the views information and semantic information, they reinforce each other in the performance gain.

5) Discussions About Image Features: Recently, using deep learning features for image classification and recognition is very popular [52]. In order to demonstrate the efficient performance of our method, we add an experiment which replaces the color and texture feature with the AlexNet feature [52], we denote this experiment as the SR-AlexNet. The performance comparisons are illustrated in Figs. 9 and 10.

Form the Figs. 9 and 10, we can see that using AlexNet feature can make the relevance performance better, and also gain some diversity improvement for the top 15 ranked results. However,

2.80 respectively, while their MADP@20 values are 1.16, 1.667, 1.08, 1.07,1.814, and 2.148 respectively.

We can see that the VR achieves a little higher NDCG, MAP much lower MADP than the VUR. From this, we can acknowledge that user information contributes to the promotion of the diversity performance. However, without the cooperation of the appropriate intra-user re-ranking, the improvement of the diversity performance is at the cost of the reduction of the relevance performance, just as Figs. 5 and 6 shown. When the intra-user re-ranking and inter-user re-ranking are combined, SR obtain higher NDCG, MAP than the VR and higher MADP than VUR.

Besides, the RR has a little lower NDCG@20, MAP@20 values and a little bigger MADP@20 value than the CRR method. But, using RR is relatively time consuming. For the RR method takes the all tags of images into consideration, CRR only considers the co-occurrence tags. Time is the other key role in the image retrieval system except the performance. So, the CRR is more suitable for the retrieval of the large database.

From the experimental results, we can find that the DRR and SR both get better diversity performance as shown in Fig. 7. However, the semantic restriction which DRR proposed to enhance the diversity performance weakens their relevance performance as shown in Fig. 5. SR makes a better trade-off between the relevance and diversity performance by considering the social user's information.

4) Discussions About Weight Selection: In this part, the impact of the regularization parameter α, β [can be found in (9)]
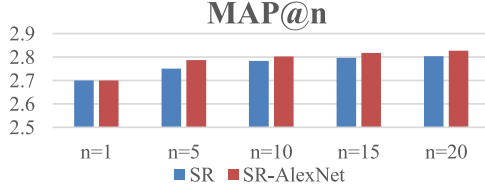
MAP@n

Fig. 9.   MAP of SR and SR-AlexNet under different depths.

MADP@n
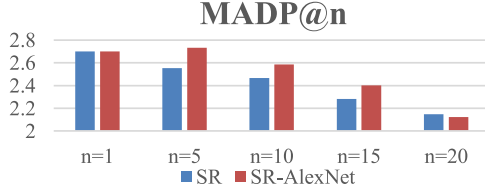

Fig. 10.   MADP of SR and SR-AlexNet under different depths.

MAP@n
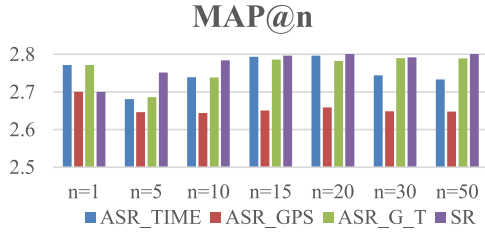

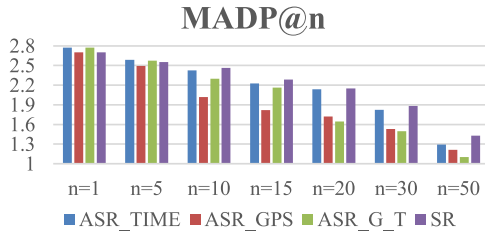Fig. 11.   Mean AP of different methods under different depths.

MADP@n

Fig. 12.   Mean ADP of different methods under different depths.

	Offline (seconds)	Online
(Seconds)		
VR	0	0.8421
VUR	0	3.8627
RR	163.2745	81.4735
CRR	38.05965	81.4735
DRR	3054.6124	83.9267
SR	38.05965	10.176

ASR_GPS: Adaptive social re-ranking based on the GPS location (geo-tag, which has been used in POI detection [47], [54] and Image Location Estimation [14]). The user's image set is grouped by mean-shift using their GPS locations, and then we select the most relevant image from each cluster in each user's image set.

ASR_GPS_TIME: Adaptive social re-ranking based on the GPS location and image taken time. In ASR_GPS_TIME, images in each season cluster, which was obtained by ASR_TIME, is grouped by mean-shift using their GPS locations with the optimal bandwidth. And then we select the most relevant image from each GPS cluster in each user's season cluster. For simplicity, we use ASR_G_T instead.

From the Fig. 11, we can see that SR achieves a higher MAP than all the ASR method when N is larger than 5. In Fig. 12, SR achieves a higher MADP than all the ASR method when N is larger than 10. We can acknowledge from Table I that the image number per user is 752. Several users only have one image under several uncommon queries. Therefore, select too many relevant images from the same user's image set will introduce irrelevant images and bring the diversity performance down. Besides, ASR_Time achieves the best performance on all the ASR method. The reason is that images taken in different seasons were seldom taken in the same places or with the same topic.

7) Computational Cost: From Section III, we can know that our proposed method SR can be divided into two parts: offline parts and online parts. In order to demonstrate the effective of SR, we illustrated the time cost of different methods per query in Table II (on a PC with intel core i5-3470 CPU and 16G memory). The following analysis don't include the extraction of visual feature, view feature, or GPS feature.

As the Section VI described, the VR and VUR don't have the offline parts, the offline parts of RR includes the process of tag filtering, the construction of the inverted index construction, visual similarity obtain and the semantic relevance obtain. The offline parts of DRR include the process of tag filtering, the construction of the inverted index construction, visual similarity obtain, the semantic relevance obtain and the pairwise tag similarity obtain. The offline parts of SR and CRR include the process of tag filtering, the construction of the inverted index construction, co-occurrence word obtain, visual similarity obtain and the semantic relevance obtain. The online parts of VR are just the view times ranking of images. VUR adds the

the 4096-dim AlexNet makes it much more complexity than our 215-dim color and texture feature, so we prefer the 215-dim color and texture feature for a better user experience.

6) Discussions About the Number of Representative Images From Each User's Image Set: In order to select a suitable representative image number from each user's image set in method SR. In this discussion process, we conducted three comparison experiments (ASR_TIME, ASR_GPS and ASR_GPS_TIME) with SR. Besides, the performance of MAP and MADP are also shown in Figs. 11 and 12.

ASR_TIME: Adaptive social re-ranking based on the image taken time. In this method, only the number of representative images which are selected from each user's image set is different from SR, other processes are all the same with SR. In ASR_TIME, the user's image set is classified into four season clusters by using the image taken time. That's to say, the images taken in spring are grouped into the spring cluster. Then we select the most relevant image from each season cluster in each user's image set. Same user's images are ranked by their relevance scores.

inter-user ranking part on VR. The online parts of RR and CRR are the relevance score obtain parts. The online parts of DRR are the relevance score obtain and diverse relevance ranking parts. The online parts of SR consist of the intra-user ranking and inter-user ranking steps.

In the computation parts of semantic similarity, SR and CRR only calculate the semantic distance between co-occurrence tag and query tag. But RR and DRR calculate all tags. Besides, the offline parts of DRR not only calculate the semantic distance between all tags and query tag, but also the semantic distance between all tags. In the online parts, SR only calculates the relevance score of per user's image dataset, while RR, CRR and DRR calculate all images' relevance scores as a whole. This computing structure of SR saves time cost in the online parts. Therefore, we can find that SR can gain the best performance at a relatively low time cost. That's to say, our proposed method can rank about 9000 images in 10 seconds.

VII. Conclusion and Future Work

In this paper, we propose a social re-ranking method for TBIR. In this social re-ranking method, inter-user re-ranking and intra-user re-ranking are carried out to obtain the retrieved results. In order to enhance the diversity performance, user information is firstly introduced into our proposed approach and obtains satisfactory results. Besides, views of social image is also firstly fused into a traditional regularization framework to enhance the relevance performance of retrieved results. Discussions and experiments have demonstrated that our proposed method is effective and time-saving.

However, in the inter-user ranking process only user's contribution is considered and the similarity among users is ignored. In addition to this, many information in Flickr dataset are still ignored, such as title information, time stamp and so on. For future work, we will investigate the similarity among user groups in Flickr dataset. Therefore, we can fuse these relationships to enhance the diversity performance of image ranking system.

References

[1] D. Liu, X. Hua, L. Yang, M. Wang, and H. Zhang, "Tag ranking," in *Proc. Int. Conf. World Wide Web*, 2009, pp. 351–360.

[2] X. Li, C. Snoek, and M. Worring, "Learning tag relevance by neighbor voting for social image retrieval," in *Proc. ACM Int. Conf. Multimedia Inform. Retrieval*, 2008, pp. 180–187.

[3] D. Liu, X. Hua, M. Wang, and H. Zhang, "Boost search relevance for tag-based social image retrieval," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jun.-Jul. 2009, pp. 1636–1639.

[4] K. Yang, M. Wang, X. Hua, and H. Zhang, "Social image search with diverse relevance ranking," in *Proc. Int. Conf. Magn. Magn. Mater.*, 2010, pp. 174–184.

[5] M. Wang, K. Yang, X. Hua, and H. Zhang, "Towards relevant and diverse search of social images," *IEEE Trans. Multimedia*, vol. 12, no. 8, pp. 829–842, Dec. 2010.

[6] A. Ksibi, A. B. Ammar, and C. B. Amar, "Adaptive diversification for tag-based social image retrieval," *Int. J. Multimedia Inform. Retrieval*, vol. 3, no. 1, pp. 29–39, 2014.

[7] Y. Gao *et al.*, "Tag-based social image search with visual-text joint hypergraph learning," in *Proc. ACM Int. Conf. Multimedia Inform. Retrieval*, 2011, pp. 1517–1520.

[8] D. Cai, X. He, Z. Li, W. Ma, and J. Wen, "Hierarchical clustering of WWW image search results using visual, textual and link information, presented at the ACM Multimedia Conf., New York, NY, USA, 2004.

[9] K. Song, Y. Tian, T. Huang, and W. Gao, "Diversifying the image retrieval results," in *Proc. ACM Multimedia Conf.*, 2006, pp. 707–710.

[10] R. Leuken, L. Garcia, X. Olivares, and R. Zwol, "Visual diversification of image search results," in *Proc. 18th Int. Conf. World Wide Web*, 2009, pp. 341–350.

[11] R. Cilibrasi and P. Vitanyi, "The Google similarity distance," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 3, pp. 1065–1076, Mar. 2007.

[12] X. Qian, H. Wang, G. Liu, and X. Hou, "HWVP: Hierarchical wavelet packet texture descriptors and their applications in scene categorization and semantic concept retrieval," *Multimedia Tools Appl.*, vol. 69, no. 3, pp. 897–920, Apr. 2014.

[13] X. Qian, G. Liu, and D. Guo, "Object categorization using hierarchical wavelet packet texture descriptors," in *Proc. IEEE 11th Int. Symp. Multimedia*, Dec. 2009, pp. 44–51.

[14] X. Qian, Y. Zhao, and J. Han, "Image location estimation by salient region matching," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4348–4358, Nov. 2015.

[15] D. Wu, J. Wu, and M. Lu, "A two-step similarity ranking scheme for image retrieval," in *Proc. IEEE 6th Int. Symp. Parallel Architectures, Algorithms Program.*, Jul. 2014, pp. 191–196.

[16] D. Zhou, O. Bousquet, T. Lal, and J. Weston. "Learning with local and global consistency," *Adv. Neural Inform. Process. Syst.*, vol. 16, no. 16, pp. 321–328, 2004.

[17] G. Agrawal and R. Chaudhary, "Relevancy tag ranking," in *Proc. IEEE 2nd Int. Conf. Comput. Commun. Technol.*, Sep. 2011, pp. 169–173.

[18] L. Chen, S. Zhu, and Z. Li., "Image retrieval via improved relevance ranking," in *Proc. IEEE 33rd Chin. Control Conf.*, Jul. 2014, pp. 4620–4625.

[19] L. Wu and R. Jin, "Tag completion for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 716–727, Mar. 2013.

[20] Y. Yang, Y. Gao, H. Zhang, and J. Shao, "Image tagging with social assistance," in *Proc. Int. Conf. Multimedia Retrieval*, 2014, Art. no. 81.

[21] L. Chen, D. Xu, and I. Tsang, "Tag-based image retrieval improved by augmented features and group-based refinement," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1057–1067, Aug. 2012.

[22] S. Lee, W. D. Neve, and Y. M. Ro, "Visually weighted neighbor voting for image tag relevance learning," *Multimedia Tools Appl.*, vol. 72, no. 2, pp. 1363–1386, 2013.

[23] Z. Xu *et al.*, "Measuring semantic relatedness between Flickr images: From a social tag based view," *Sci. World J.*, vol. 2014, no. 2014, pp. 171–192, 2014.

[24] X. Li, "Tag relevance fusion for social image retrieval," *Multimedia Syst.*, pp. 1–12, 2014.

[25] X. Qian, X. Liu, and C. Zheng, "Tagging photos using users' vocabularies," *Neurocomputing*, vol. 111, no. 111, pp. 144–153, 2013.

[26] D. Mishra, U. P. Singh, and V. Richhariya, "Tag relevance for social image retrieval in accordance with neighbor voting algorithm," *Int. J. Comput. Sci. Netw. Secur.*, vol. 14, no. 7, pp. 50–57, 2014.

[27] Y. Hu and M. Li. "Multiple-instance ranking: Learning to rank images for image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2008, pp. 1–8.

[28] F. Sun, M. Wang, and D. Wang, "Optimizing social image search with multiple criteria: Relevance, diversity, and typicality," *Neurocomputing*, vol. 95, pp. 40–47, 2012.

[29] B. Wang, Z. Li, and M. Li., "Large-scale duplicate detection for web image search," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2006, pp. 353–356.

[30] K. Weinberger and M. Slaney, "Resolving tag ambiguity," in *Proc. 16th ACM Int. Conf. Multimedia*, 2008, pp. 111–120.

[31] A. Ksibi, G. Feki, A. B. Ammar, and C. B. Amar, "Effective diversification for ambiguous queries in social image retrieval," in *Computer Analysis of Images and Patterns*. Berlin, Germany: Springer, 2013, pp. 571–578.

[32] A. Khosla, A. Sarma *et al.*, "What makes an image popular?," in *Proc. Int. Conf. World Wide Web*, 2014, pp. 867–876

[33] C. Haruechaiyasak, *Improving Social Tag-Based Image Retrieval With CBIR Technique*. Berlin, Germany: Springer, 2010, pp. 212–215.

[34] X. Zhu and W. Nejdl, "An adaptive teleportation random walk model for learning social tag relevance," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, 2014, pp. 223–232.

[35] J. Yu, D. Tao, M. Wang, and Y. Rui, "Learning to rank using user clicks and visual features for image retrieval," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 767–779, Apr. 2014.

[36] S. Ji, K. Zhou, C. Liao, Z. Zheng, and G. R. Xue, "Global ranking by exploiting user clicks," in *Proc. 32nd Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, 2009, pp. 35–42.

[37] G. Dupret and C. Liao, "A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine," in *Proc. 3rd ACM Int. Conf. Web Search Data Mining*, 2010, pp. 181–190.

[38] S. Cen, L. Wang, Y. Feng, and H. Bai, "Efficient image reranking by leveraging click data," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops*, Jul. 2014, pp. 1–4.

[39] X. Hua and M. Ye., "Mining knowledge from clicks: MSR-Bing image retrieval challenge," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops*, Jul. 2014, pp. 1–4.

[40] L. Chen, D. Xu, and I. Tsang, "Tag-based web photo retrieval improved by batch mode re-tagging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 3440–3446.

[41] A. Sun and S. Bhowmick, "Image tag clarity: In search of visual representative tags for social images," in *Proc. 1st SIGMM Workshop Social Media*, 2009, pp. 19–26.

[42] X. Qian, X. Hua, Y. Tang, and T. Mei, "Social image tagging with diverse semantics," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2493–2508, Dec. 2014.

[43] X. Qian, D. Lu, and X. Liu, "Tag based image retrieval by user-oriented ranking," in *Proc. 5th ACM Int. Conf. Multimedia Retrieval*, 2015, pp. 511–514.

[44] K. Jarvelin and J. Kekalainen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inform. Syst.*, vol. 20, no. 4, pp. 422–446, Oct. 2002.

[45] Y. Gu, X. Qian, and Q. Li., "Image annotation by latent community detection and multikernel learning," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3450–3463, Nov. 2015.

[46] X. Yang, X. Qian, and Y. Xue, "Scalable mobile image retrieval by exploring contextual saliency," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1709–1721, Jun. 2015.

[47] S. Jiang, X. Qian, J. Shen, Y. Fu, and T. Mei, "Author topic model-based collaborative filtering for personalized POI recommendations," *IEEE Trans. Multimedia*, vol. 17, no. 6, pp. 907–918, Jun. 2015.

[48] X. Qian, Y. Xue, Y. Tang, and X. Hou, "Landmark summarization with diverse viewpoints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1857–1869, Nov. 2015.

[49] D. Liu, G. Ye, C. T. Chen, S. Yan, and S.-F. Chang, "Hybrid social media network," in *Proc. 20th ACM Int. Conf. Multimedia*, 2012, pp. 659–668.

[50] G.-J. Qi, C. C. Aggarwal, J. Han, and T. Huang, "Mining collective intelligence in diverse groups," in *Proc. 22nd Int. Conf. World Wide Web*, 2013, pp. 1041–1052.

[51] G.-J. Qi, C. C. Aggarwal, Q. Tian, H. Ji, and T. Huang, "Exploring context and content links in social media: A latent space method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 850–862, May 2012.

[52] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inform. Process. Syst. 25*, 2012, pp. 1106–1114.

[53] S. Wang, Q. Huang, S. Jiang, and Q. Tian, "S3MKL: Scalable semi-supervised multiple kernel learning for real world image data mining," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1259–1274, Aug. 2012.

[54] J. Zhang, S. Wang, and Q. Huang, "Location-based parallel tag completion for geo-tagged social photo retrieval," in *Proc. Int. Conf. Multimedia Retrieval*, Jun. 2015, pp. 355–362.

[55] X. Lei, X. Qian, and G. Zhao, "Rating prediction based on social sentiment from textual reviews," *IEEE Trans. Multimedia*, to be published.

[56] G. Zhao, X. Qian, and X. Xie, "User-service rating prediction by exploring social users' rating behaviors," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 496–506, Mar. 2016.

[57] X. Qian, X. Liu, X. Ma, D. Lu, and C. Xu, "What is happening in the video? Annotate video by sentence," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.

[58] X. Qian, X. Tan, Y. Zhang, R. Hong, and M. Wang, "Enhancing sketch-based image retrieval by re-ranking and relevance feedback," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 195–208, Jan. 2016.

[59] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized recommendation combining user interest and social circle," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp. 1487–1502, Jul. 2014.

Dan Lu received the B.S. degree from Chang'an University, Xi'an, China, in 2013, and is currently working toward the M.S. degree at the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, China.

She is now a Post-Graduate at the SMILES Laboratory, Xi'an Jiaotong Univeristy. Her current research interests include tag-based image retrieval.

Xiaoxiao Liu received the B.S. degree from the Xi'an University of Post and Telecomunications, Xi'an, China, in 2011, and the M.S. degree from the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, China, in 2014.

Xueming Qian (M'09) received the B.S. and M.S. degrees from the Xi'an University of Technology, Xi'an, China, in 1999 and 2004, respectively, and the Ph.D. degree from the School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, China, in 2008.

He was a Visiting Scholar with Microsoft Research Asia, Beijing, China, from 2010 to 2011. He was previously an Assistant Professor with Xi'an Jiaotong University, where he was an Associate Professor from 2011 to 2014, and is currently a Full Professor. He is also the Director of the Smiles Laboratory, Xi'an Jiaotong University. His research interests include social media big data mining and search.

Prof. Qian was the recipient of the Microsoft Fellowship in 2006. He was also the recipient of outstanding doctoral dissertations of Xi'an Jiaotong University and Shaanxi Province, in 2010 and 2011, respectively.