Image Retrieval by User-oriented Ranking

Xueming Qian
Xi'anJiaotong University,China
qianxm@mail.xjtu.edu.cn

Dan Lu
Xi'anJiaotong University,China
ludandan@stu.xjtu.edu.cn

Xiaoxiao Liu
Xi'anJiaotong University,China
xiaoxiaoliu@stu.xjtu.edu.cn

ABSTRACT
Tag-based image search is an important method to process images contributed by social users in social media sharing websites like Flickr. However, existing ranking methods for tag-based image search frequently return results that are irrelevant, low-diversity or time-consuming. In this paper, we propose a user-oriented image ranking system with the consideration of image relevance, diversity and computation complexity, aiming to automatically rank images according to their visual information, semantic information and social clues. When you input a query in the user-oriented image search engine, images tagged with query are obtained as the initial results. The initial results include images contributed by different social users. Usually each user contributes several images. First we sort these users by inter-user ranking. Users that have a higher contribution to the given query rank higher. Then we sequentially implement intra-user ranking on the ranked user's image set, and only the most relevant image in each user's image set is selected. These selected images compose the final retrieval results. Experimental results on Flickr dataset show that our user-oriented ranking method is effective and efficient.

Categories and Subject Descriptors
H.3.3 **[information system]:** information search and retrieval – *Retrieval models*

General Terms
Algorithms, Measurement, Experimentation.

Keywords
Social Media; Tag-based Image Retrieval; Social Clues; Co-occurrence word.

1. INTRODUCTION
With the development of social media and web 2.0, massive amounts of images and videos spring up everywhere on the Internet. This phenomenon has brought enormous challenges to multimedia storage, indexing and retrieval. Generally speaking, tag-based image search is the most commonly used image retrieval method in social media, which enables users to formulate their queries using tags.

Nonetheless, users cannot precisely describe their request with single tags. Therefore, the query ambiguity problem comes into play.Thus, a fundamental problem in social image retrieval is how to reliably solve "query ambiguity" problems. To solve the "query ambiguity" problem, an effective approach is to provide diverse retrieval results that cover multiple topics underlying a query. Currently, image clustering [1, 3] and duplicate removal [5,6,8,9] are the major approaches in settling the diversity problem. In [1], Cai et al. proposed a hierarchical clustering method to cluster the search results into different semantic clusters by a multimodal fusion method, which fuses visual, textual and link analysis. Similarly, In [4], Leuken et al. studied three visually diverse ranking methods to re-rank the image retrieval results, through clustering the visual features of the images. Different from clustering, Ksibi et al. [9] proposed an adaptive diverse relevance ranking algorithm to diversify the relevant search results for ambiguous queries in tag-based social image retrieval using an adaptive diversification trade-off. Yang and Wang et al. [5-6] 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. [8] proposed a DTRR method to promote the diversity performance of retrieval results by fusing both semantic and visual information of images on the basis of [6]. Despite satisfactory results have been obtained, most of them highly rely on the visual and semantic information, and ignore the essence of social images. Social images uploaded and tagged by users are user-oriented. These images are taken in user favorite angle at user available time. For the same scenic spots, various users show us diverse visual effects. Thus, user information cannot be neglected in enhancing the diversity of the retrieval results. Motivated by this intuition and above analysis, we propose a user-oriented image ranking algorithm which introduces the user information into the traditional ranking method. The contributions of this paper can be described as follows: 1)We propose a user-oriented ranking method to enhance the diversity performance, which is complementary to the existing ranking methods.2)We propose a novel algorithm to obtain the co-occurrence word set of the given query which can be utilized to boost the relevance level of the tag-based image retrieval and reduce the computational cost of our user-oriented image ranking algorithm.3)We introduce a regularization framework to evaluate the relevance scores of images with respect to the given query. Visual feature and semantic feature are merged into this framework.

The remainder of the paper is organized as follows. The system overview is illustrated in section 2. Section 3 demonstrates the offline system. The online system is depicted in section 4. Experiments on Flickr dataset are setup and shown in section 5. Finally, the conclusion and future work are given in section 6.

2. SYSTEM OVERVIEW
Our user-oriented image ranking system includes two main sections: Online and Offline. The whole system is shown in Figure 1. The offline section is the feature extraction section. In this paper, we extract the visual feature and semantic features of our image dataset. Semantic features refers to the co-occurrence word set of the given query, and tags of the images. Our online section consists of the following 2 steps: 1) Inter-user ranking. The inter-user ranking is applied to rank the corresponding users with the consideration of their contribution to the given query. 2) Intra-user ranking. A regularization framework is included in this process to attach a relevance score to each image of each user which measures their relevance level of the given query. Then, the most relevant image

ICMR'15, June 23–26, 2015, Shanghai, China.
@ 2015 ACM. ISBN 978-1-4503-3274-3/15/06 …$15.00.
DOI:http://dx.doi.org/10.1145/2671188.2749326

is sequentially selected on the image dataset of each ranked user, which consists of our retrieval results. Hereinafter the details are displayed.

3. THE OFFLINE SYSTEM

In our offline system, the feature extraction of the image database is implemented. In this paper, we use the visual features and semantic features to represent our image dataset. The details are as follows.

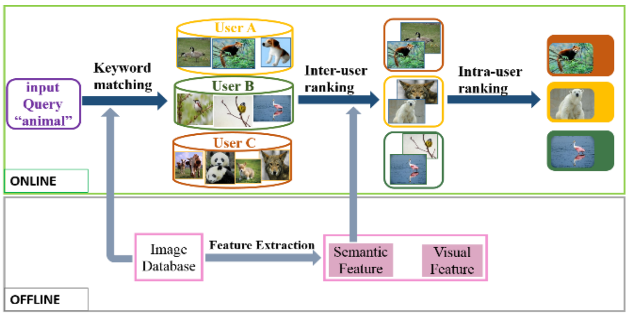


Figure 1. User-oriented image ranking system.

3.1 Visual Features

In this paper, a 215-dimensional visual vector is utilized, including a 45-dimensional color moment feature, and a 170-dimensional texture feature vector [2], [7].

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 the Gaussian kernel function with a radius parameter σ [3].

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

where $\|.\|^2$ stands for the l2-norm of the vector. Furthermore, σ represents the radius parameter set to be the mean value of all pairwise Euclidean distance between images.

3.2 Semantic Features

In this paper, the semantic features of image is its associated tags; the semantic features of the given query is its co-occurrence word set. Co-occurrence means two words which often appeared in the text corpus in a certain order. It can also be interpreted as an indicator of semantic proximity or an idiomatic expression. Suppose that $E(q)=\{e_1,e_2,\ldots,e_l\}$ denote the co-occurrence word set about query q, l is the number of co-occurrence word with respect to the query q. Then, the image which tagged with $e_i \in E(q)$ and q is much more relevant with query q than the image which is not. Thus, we can obtain the co-occurrence word set $E(q)=\{e_1,e_2,\ldots,e_l\}$ as follows:

1) For the tag q, we get its top P tags according to their frequencies in the search results which were issued by query q in our image dataset. In this paper, we set P=100 (q is not included in the top P tags).
2) We select the tags which satisfy the following rules in the tag set selected by step 1), If $\frac{R(q,a)}{R(q)} > d*R(a)/N$, where R(q, a) is the number of images which tagged with tag q and tag a in Flicker. R(q) is the number of images which tagged with tag q in Flicker; N demonstrates the total number of images in Flickr. Then the tag a is selected, and we arranged these selected tags by R(q, a) in a descending order, which denotes S. The parameter d set to be 150 in our paper.
3) We choose the top v tags in the selected tag set S as the co-occurrence word set, where v satisfies the rule: the difference between the v-th tag frequency given the query and the (v+1)-th tag frequency given the query is maximum in tag set S.

The co-occurrence words we obtained are: sky, sun and cloud; coast, sand, ocean, and sea; band and concert; airplane, airport and aircraft; and so on. However, each element in E(q) has a different degree of importance in boosting the relevance score of the retrieval results. 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. After obtaining the co-occurrence word set E(q) for each query q, we assign them weights which are denoted by $M(E(q))=\{M_1,M_2,\ldots,M_l\}$ based on their co-occurrence similarity, which is determined as follows:

$$M_i = exp\left\{-\frac{\max\{\log R(q),\log R(e_i)\} - log\ R(q,e_i)}{log\ N - \min\{\log R(q),\log R(e_i)\}}\right\} \tag{2}$$

Where $R(q)$, $R(q,e_i)$, N have already been defined above.

It's acknowledged that the tags associated with an image are arranged in a random order without any important or relevant information, which limits the effectiveness of tag-based image retrieval. So, we need to measure the semantic relevance level between the tag and image. Thus, a semantic relevance matrix C is put forward to measure the semantic relevance between query tagged image and the query tag. We define the average co-occurrence similarity between the query and the tag set of image i as C_i ,which is calculated as follows:

$$C_i = \frac{1}{\sum_{e_m \in E(q)} 1(e_m)} \sum_{e_m \in E(q)} M_m * 1(e_m) \tag{3}$$

where $1(e_m)$ denotes whether the image *i* containing tag e_m or not, i.e.

$$1(e_m) = \begin{cases} 1; & \text{if image i is tagged with } e_m \\ 0; & otherwise \end{cases} \tag{4}$$

4. THE ONLINE SYSTEM

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

For the query q, we can obtain the corresponding images that all tagged with query q in our dataset, which is denoted by X. The following steps are all carried out in dataset X. X also can be further described as the follows by taking the social user's information into account, i.e.

$$X = \{X(u_1),\ldots,X(u_Z)\} = \{X_1,\ldots,X_Z\} = \\ \{\{x_{11},x_{12,}\ldots,x_{1N_1}\},\{x_{21},x_{22,}\ldots,x_{2N_2}\},\ldots,\{x_{Z1},x_{Z2,}\ldots,x_{ZN_Z}\}\} \tag{5}$$

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 in X uploaded by i-th user u_i ; x_{ij} is the j-th image in image dataset X_i ; N_i denotes the number of images in X_i .

4.1 Inter-user Ranking

Inter-user ranking ranks the users who uploaded images in X according to their contribution to the query q. Larger contribution users probably show viewers more professional images. And this contribution is measured upon the number of its images in X which are tagged with words in E(q).

For each user U_h, $h \in (1,2,\ldots,Z)$, we calculate its contribution to the query (denoted by UW_h) as follows:

$$UW_h = \sum_{j=1}^{k} 1(x_{hj}) \tag{6}$$

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

$$1(x_{hj}) = \begin{cases} 1; & t_{hj} \cap E(q) \neq \varnothing \\ 0; & otherwise \end{cases} \tag{7}$$

In the above Equation, t_{hj} is the tag set of image x_{hj}.

We rank $UW_h, \mathrm{h} \in (1,2,\ldots,\mathrm{Z})$ in a descending order. The larger UW_h, corresponding user ranks higher.

4.2 Intra-user Ranking

After inter-user ranking, the largest contribution user ranks highest. Then we implement intra-user ranking to select the most relevant image among each user's image set. We take the image set $X_h, \mathrm{h} \in (1,2,3,\ldots,\mathrm{Z})$ as an example to demonstrate our intra-user ranking process.

For each image in $X_h = \{x_{\mathrm{h}1}, x_{\mathrm{h}2}, \ldots, x_{hk}\}$, its relevance score is $r = [r_1, r_2, ..., r_k]$. We introduced a regularization framework to obtain r, which takes the visual and semantic information into account. The regularization framework [3] is defined as follows:

$$\mathrm{Q}(\mathrm{r}) = \sum_{i=1}^{k}\sum_{j=1}^{k} w_{ij}\left\{r_i / \sqrt{D_{ii}} - r_j / \sqrt{D_{jj}}\right\}^2 + \lambda \sum_{i=1}^{k}(r_i - C_i)^2 ? \tag{8}$$

where Q(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. C_i is the semantic relevance score of image i with respect to the query q. The first term in the right-hand side of the cost function means that the relevance scores of visually similar images should be close, and the second term is a fitting constraint, which means that the relevance scores are biased with preference to the semantic relevance measurement, $\lambda > 0$ is the regularization parameter, which denotes the trade-off between these two competing items.

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

$$r^* = argmin(Q(r)) \tag{9}$$

To address the optimization problem (9), Equation (8) can be rewrite as the matrix form as:

$$\mathrm{Q}(\mathrm{r}) = r^T\left(I - D^{-\frac{1}{2}}WD^{-\frac{1}{2}}\right)r + \lambda\|r - C\|^2 \tag{10}$$

where $D = Diag(\mathrm{D}_{11}, \mathrm{D}_{22}, \ldots, \mathrm{D}_{kk})$, $C = [C_1, C_2, \ldots, C_k]$.

Alternatively, we can use iterative optimization algorithm to solve this problem, which avoids the intensive computation brought by the direct matrix inversion in Equation (10). The detailed steps are as follows: (1) utilize the image affinity matrix W determined by Eq.1, if $i \neq j$ and otherwise $\mathrm{w}_{ii} = 0$. (2) compute the semantic relevance matrix C, which is illustrated in section 3.2. (3)iterate $\mathrm{r}(\mathrm{t}+1) = \frac{1}{1+\lambda}D^{-\frac{1}{2}}WD^{-\frac{1}{2}}\mathrm{r}(\mathrm{t}) + \frac{\lambda}{1+\lambda}C$ until convergence, then the optimization relevance score r of each image in X_h can be achieved.

After obtaining the relevance score r of each image in X_h, $h = \{1,2,...,Z\}$, we select the image with the highest score, which is denoted by x_{fh}, $h = \{1,2,...,Z\}$. Finally, we arrange the image set by $\{x_{\mathrm{f}1}, x_{\mathrm{f}2}, \ldots, x_{fZ}\}$ the order of their users which was arranged by their contributions. These ranked images constitute our retrieval results.

5. EXPERIMENTS

In order to demonstrate the effectiveness of the user-oriented ranking approach, we conduct an experimental comparison between the following 4 methods on 20 popular tags: airplane, beach, Beijing, bird, blue, buildings, Christmas, cityscape, forest, reflection, garden, girls, honeybee, insect, lotus, ocean, orange, sea, sky, Zebra. Our initial Flickr dataset includes 6600034 images uploaded by 7249 users and their related files recoding the information of tags and users. We remove the images that have no tag information, then 5325265 images and 7090 users left.

- Relevance ranking [3], an optimization framework is applied to automatically rank images based on visual and semantic information of images. For simplicity, we use the RR instead.
- Co-occurrence relevance ranking: an algorithm which is the same with RR except that the semantic relevance score in RR is replaced by the semantic relevance score of our proposed approach. For simplicity, we use the CRR instead.
- Diverse relevance ranking [6], which optimizes an ADP measure with the consideration of the semantic and visual information of images. We use DRR for simplicity.
- User-oriented ranking. The proposed method which fuses the user information into CRR method. We use the UOR for simplicity.

5.1 Performance Evaluation

In this part, our experiments are performed to demonstrate whether we have selected a better method for tag-based image retrieval. And the parameter λ in (10) is empirically set to be 0.1 for all queries.

The performance evaluation of our method is voted by three volunteers and this evaluation contains two parts: relevance score and diversity score. The relevance score is used to evaluate the correlation between the query and the retrieval results. And the diversity score shows the diversity performance of the retrieval results.We voted on the relevance and diversity score of the top-n results obtained through four different methods. The relevance score of the i-th image is $\mathrm{rel_i}, \mathrm{i} = 1,2,\ldots,\mathrm{n}$, Each image is labeled as one of four levels: excellent(score:3), good(score:2), so-so(score:1), irrelevant (score:0). Their overall diversity score is div@n, which is labeled as one of four levels: excellent(score:3),good(score:2),soso(score:1),similar(score:0).

Once we get the value of $\mathrm{rel_i}$ and div@n, the ADP of the top-n image can be obtained from equation (11), (12) and (13).

$$AP@n = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{i}\frac{rel_j}{i}\right) \tag{11}$$

$$ADP@n = AP@n * norm\_div@n \tag{12}$$

$$norm\_div@n = \frac{div@n}{3} \tag{13}$$

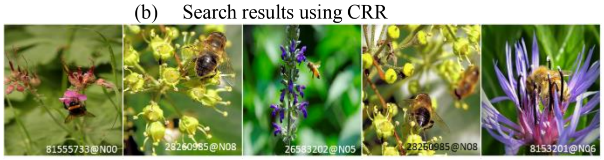
5.2 Experimental Results

The top-5 results of exemplar query "honeybee" on Flickr database under the four different ranking algorithms are illustrated in Figure

2, from which we can see that the results of UOR are both relevant and diverse, whereas the others suffer from the lack of diversity. And most of the similar images in the top-5 results of RR, CRR ,DRR are comes from the same user. From this, the effectiveness of our user-oriented idea has been proven.

(a) Search results using RR

(b) Search results using CRR



(c) Search results using DRR

(d) Search results using UOR

Figure.2. **Top 5 ranking results of an exemplar query "honeybee".**

5.3 Performance Analysis

In this section, we compute ADP@20 values of all the 20 tags based on the voting. Table1 shows the mean ADP@20 and mean div@20 values of all 20 queries under the four different methods. The MAP@n values of all these four models are shown in Figure 3 to evaluate their relevance performance, where 6 different depths n are displayed for a clear see.

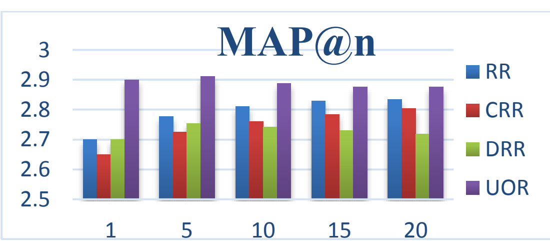


Figure.3 The Mean AP@n of all 4 ranking methods under different depths.

Table 1. The overall performance of each 4 ranking results

METHOD	RR	CRR	DRR	UOR
MADP@20	1.065	1.283	1.952	2.309
Mdiv@20	1.165	1.365	2.17	2.4

From the above charts, we can discover UOR achieves the highest scores on MAP, Mdiv@20 and mean MADP@20, which demonstrate UOR outperforms other methods. The RR has a slighter bigger MAP value and a much lower MADP@20 than the CRR method. Besides, 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. When we fuse the user information into the CRR method, its MADP@20 is larger than the DRR method, diversity performance is bigger than DRR. For the semantic similarity constrain which the DRR proposed degrades the relevance degree of the results. While UOR is able to achieve a good trade-off between the relevance and diversity performance of the retrieval results. Furthermore, introducing user information adds little computation complexity to UOR. This proves the UOR is appropriate for conducting the retrieval of the large image database.

In summary, the comparison experiment results confirm the proposed method from the following three aspects: (1) The MADP@20 of the retrieval results can be promoted significantly through UOR, which can provide a rich related information to the user. (2) The semantic relevance estimation method which UOR proposed by using co-occurrence word set is not only timesaving, but also effective in measuring the semantic relevance of images with respect to the given query. (3) The computational cost of UOR method is low.

6. CONCLUSIONS

In this paper, we propose a user-oriented image ranking method for tag-based image retrieval. It leverages the visual, semantic information and social clues of images. In order to accomplish the ranking, inter-user ranking and intra-user ranking are carried out to obtain the search results. The experimental results demonstrate that our user-oriented image ranking is effective.

Acknowledgments: This work is supported in part by the Program 973 No. 2012CB316400, by NSFC No.60903121, 61173109, 61202180, 61332018, Microsoft Research Asia, and Fundamental Research Funds for the Central Universities.

7. REFERENCES

[1] D. Cai, X. He, Z. Li, WY. Ma, JR. Wen. "Hierarchical clustering of WWW image search results using visual, textual and link information," In Proceedings of the 12th annual ACM international conference on Multimedia, pp. 952-959, 2004.

[2] X. Qian, G. Liu, D. Guo. Object categorization using hierarchical wavelet packet texture descriptors. in Proc. ISM 2009, pp.44-51.

[3] D. Liu, XS. Hua, M. Wang. "Boost search relevance for tag-based social image retrieval," In *IEEE* International Conference, pp. 1636-1639, 2009.

[4] .RH. van Leuken, L. Garcia, X. Olivares. "Visual diversification of image search results." In Proceedings of the 18th international conference, pp. 341-350, ACM, 2009.

[5] K. Yang, M. Wang, XS. Hua. "Social image search with diverse relevance ranking." In Multimedia Modeling, pp. 174-184, 2010.

[6] M. Wang, K. Yang, XS. Hua. "Towards a relevant and diverse search of social images." Multimedia, IEEE Transactions on,12(8), 829-842, 2010.

[7] X. Qian, H. Wang, G. Liu, X. Hou, "HWVP: Hierarchical Wavelet Packet Texture Descriptors and Their Applications in Scene Categorization and Semantic Concept Retrieval", Multimedia Tools and Applications, May 2012.

[8] F. Sun, M. Wang, D. Wang. "Optimizing social image search with multiple criteria: Relevance, diversity, and typicality." Neurocomputing,95, 40-47, 2012.

[9] A. Ksibi, AB. Ammar. "Adaptive diversification for tag-based social image retrieval." International Journal of Multimedia Information Retrieval, 3(1), 29-39, 2014.