Exploiting User-supplied Tags for Flickr Photos Annotation

Zheng Liu[1,2] , Hua Yan[1,2]
1 School of Computer Science and Technology, Shandong Economic University, Ji'nan 250014, China
2 Shandong Provincial Key Laboratory of Digital Media Technology, Ji'nan 250014, China
Email: Lzh_48@126.com

***Abstract*—The popularity of photo-sharing websites like Flickr give us a chance to observe what ordinary users do in their daily life. Particularly, Flickr allows the users to provide personalized tags when uploading photos, and then we can annotate Flickr photos using user-supplied tags. This paper proposes an approach to automatically annotate Flickr photos by exploiting user-supplied tags. User-supplied tags are submitted to Wikipedia to prune noisy tags, and then the reserved tags are denoted as initial tags. Afterwards, the initial tags are ranked using manifold-ranking algorithm, by which regions of the photo to be annotated are served as queries to launch the manifold-ranking algorithm which ranks the initial tags according to their relevance to the queries. Next, using Flickr API, top ranked initial annotations are expanded by a weighted voting scheme. Finally, we combine top ranked initial tags with expanding tags to construct final annotations. Experiments conducted on Flickr photos show the effectiveness of the proposed approach.**

***Index Terms*—Manifold-ranking, Flickr Photos Annotation, SIFT, Locality-Sensitive Hashing**

I. Introduction

In recent years, we have witnessed an explosion of Web photo community site, such as Flickr, which enables users to upload and share personal photos. Such social photo repositories allow users to upload personal photos and annotate content with descriptive keywords which is called tags. With the rich tags as metadata, users can more conveniently organize and access shared photos. Making full use of the tags provided by users, a high efficient method can be proposed to annotate Flickr photos.

With the rapid development of Web social community, the applications which exploit the social media resources, such as Flickr and Wikipedia, have become popular and attracted much attention from both academia and industry [3]. Many recent Flickr-based research efforts explore correlations between keywords derived from these resources to extract image semantics. The tags which describe the content of images can help users easily manage and access large-scale image datasets. With these metadata, the manipulations of image data can be easier to be accomplished, such as browsing, indexing and retrieval [1].

Several pioneering works related to Flickr tags have been proposed, which mine useful information from Flickr photos and tags. Liu et al. proposed an approach to rank the tags for each image according to their relevance levels [1]. A new Flickr distance was proposed to measure the visual similarity between concepts according to Flickr [4]. Schmitz proposed the building of facted ontology from Flickr' tagging resources [5]. Chen et al. also proposed to use the predicted tags to search for groups as recommendation groups for the given image, however this method heavily relies on the performance of tag prediction [6]. The learning based tag recommendation approach has been introduced to generate ranking features from multi-modality correlations, and learns an optimal combination of these ranking features by the Rankboost algorithm [7]. Ames et al. have explored the motivation of tagging in Flickr website and they claim that most users tag images to make them better accessible to the general public [8]. Kennedy et al. have evaluated the performance of the classifiers trained with Flickr images and associated tags and demonstrate that tags provided by Flickr users actually contain many noises [9].

This paper presents a novel approach to automatically annotate Flickr photos with tag ranking and tag expanding. The rest of the paper is organized as follows. Section 2 introduces the framework of our Flickr photo annotation method. Manifold-ranking based initial user-supplied tag ranking and top ranked initial tags expanding are described in section 3 and 4 respectively. In section 5, the experimental results demonstrate the performance of our photo annotation method. Section 6 concludes the whole paper and points out our future works.

II. Overview of Our Methods

The source data for the annotating process is user-supplied tags. After initial tags ranking and tags expanding, the final annotations can be obtained. Our Flickr photo annotation method is made up of three steps, which is shown in Fig.1. An example of a Flickr photo with user-supplied tags is described in Table 1.

doi:10.4304/jsw.6.10.1898-1905

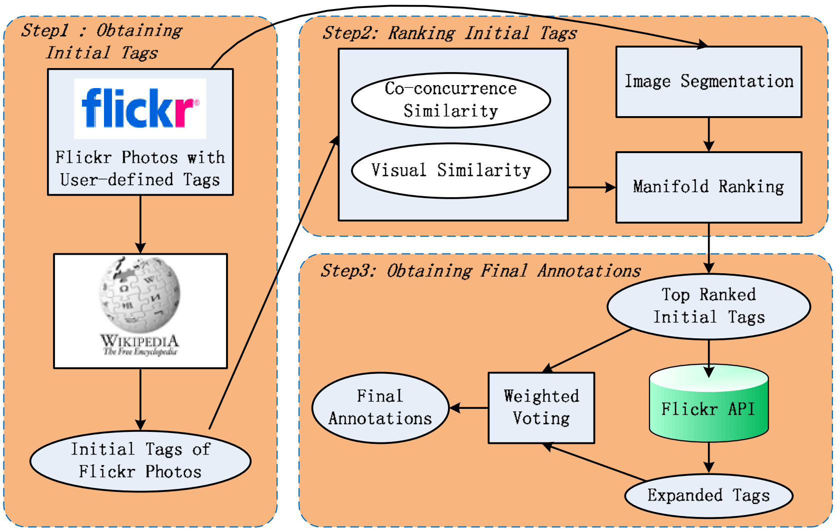
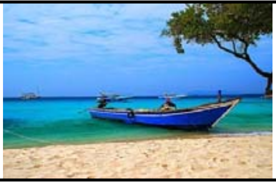

Step1 : Obtaining Initial Tags
flickr
Flickr Photos with User-defined Tags
WIKIPEDIA
The Free Encyclopedia
Initial Tags of Flickr Photos
Step2: Ranking Initial Tags
Co-concurrence Similarity
Visual Similarity
Image Segmentation
Manifold Ranking
Step3: Obtaining Final Annotations
Top Ranked Initial Tags
Flickr API
Weighted Voting
Final Annotations
Expanded Tags

Figure 1. Framework of our approach.

TABLE I. EXAMPLE OF A FLICKR PHOTO WITH USER-SUPPLIED TAGS

URL	http://www.flickr.com/photos/49206401@N00/2195052960/
Photo in Flickr	
User-supplied tags	Coral, beach, Thailand, koh larn, Pattaya, boat, sand, blue, Nikon D40x, Corals island, sea, seascape, vacation, D40x, Nikon, PhotoFaceOffWinner

III. INITIAL TAGS RANKING

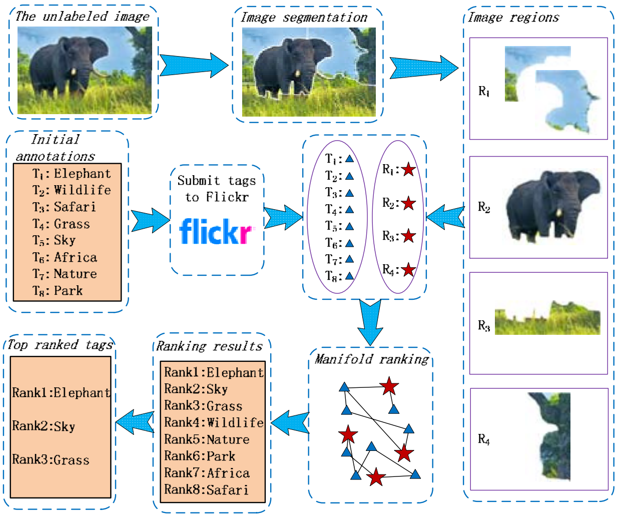


Figure 2. Illustration of tag ranking process.

The user-supplied tags may contain noisy or uncorrelated tags, such as misspelling, meaningless words and numbers. Therefore, we should perform a pre-processing method to prune the un-related tags. We submit each tag as a query to Wikipedia, and only the tags which have a coordinate in Wikipedia are reserved. After the un-related tags pruning, the rest of user-supplied is denoted as initial tags(denoted as Γ). In this section, we will show how to rank the initial tags by manifold-ranking algorithm. The tag ranking framework is shown in Fig.2.

A. Tag Similarity Measuring Scheme

Traditional image annotation methods mainly apply semantic similarity only to estimate the relevance between two words. However, it is not suitable in visual domain application such as image annotation, as this method does not take image content into account. In this paper, we design a two-level scheme considering both image content and semantic relationship to generate tag similarity.

We define a method named NFD which is analogous to NGD[15] to compute the concurrence similarity between tags based on their concurrence. NGD is a distance function between two words obtained by searching a pair of words using the Google search engine. As is shown in Fig.3, NFD between two tags can be estimated based on Flickr as follows.

$$NFD(t_i,t_j)=\frac{\max\left\{\log f(t_i),\log f(t_j)\right\}-\log f(t_i,t_j)}{\log G-\min\left\{\log f(t_i),\log f(t_j)\right\}} \qquad (1)$$

where t_i and t_j represent the two tags in consideration. $f(t_i)$ and $f(t_j)$ are the numbers of images containing tag t_i and tag t_j respectively, which can be obtained by performing search by tag on Flickr website using the tags as keywords. $f(t_i, t_j)$ is the number of the images returned by Flickr when typing t_i and t_j as the search term respectively. Moreover, G is the total number of images in Flickr. The concurrence similarity between tag t_i and tag t_j is then defined as.

$$\gamma_s(t_i, t_j) = \exp[-NFD(t_i, t_j)] \qquad (2)$$

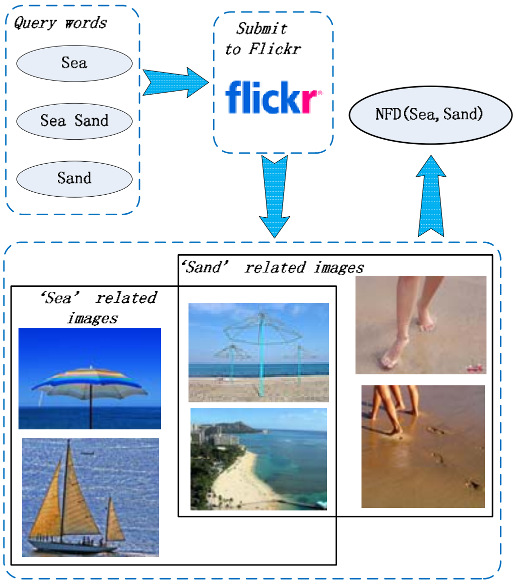


Figure 3. Illustration of NFD distance computing.

To deal with the large number of local features extracted from the images, we employ locality-sensitive hashing (LSH) algorithm which is an approximate kNN technique introduced by Indyk et al.[18] to index the local descriptors. Consider any LSH family H. The algorithm has two main parameters: the width parameter k and the number of hash tables L. In the first step, we define a new family G of hash functions g, where each function set g is obtained by concatenating k functions $h_1, h_2, \cdots, h_k$ from family G. The algorithm then constructs L hash tables, each corresponding to a different randomly chosen hash function g.

In our work, the hash functions which have been designed by Datar et al.[19] is used to construct the LSH family. Formally, for a fixed a and b, the hash function $h_{a,b}(v)$ is defined as.

$$h_{a,b}(V) = \left\lfloor \frac{a \cdot V + b}{r} \right\rfloor \qquad (3)$$

where a is a d-dimensional random vector with entries chosen independently from a Gaussian distribution and b is a real number chosen uniformly from the range $[0, r]$. r defines the quantization of the features and V is the original feature vector. For a pair of images I_i and I_j, our local feature based image similarity measuring method can be summarized as follows.

1. Extracting the SIFT descriptors from I_i and I_j.

2. Constructing L LSH hash tables, and each hash table is corresponding to a set of hash functions which is defined in Eq.3.

3. For each hash table ht_l ($1 \le l \le L$), k hash functions of which are represented as $g_l = \{h_1^l, h_2^l, \cdots, h_k^l\}$.

4. For a hash table generated from step 2, SIFT descriptors are mapped to buckets by hash functions.

5. Following step 4, for a hash table, a histogram is obtained by the way that each bin is corresponding to a bucket of the hash table.

Furthermore, local features based visual similarity of a pair of images is computed by estimating the distance between histograms. Particularly, we denote the histogram set of image I_i as $HG(i) = \{hg(i)^l \,|1 \le l \le L\}$, where $hg(i)^l$ is the l-th histogram of image I_i.

Afterwards, We follow Sung-Hyuk Cha's work[20] to compute histogram-based image visual similarity. An image could be represented by a set of SIFT descriptors, that is, $I = \{s_1, s_2, \cdots, s_n\}$. Let b be a measurement, or feature, which can have one of m values contained in the set, $B = \{b_0, b_1, \cdots b_m\}$ and $s_i \in B$ is satisfied. Hence, the corresponding histogram of image I is denoted as $hg(I) = [hg_0(I), hg_1(I), \cdots, hg_{m-1}(I)]$, where $hg_j(I)$ is the number of elements which is allocated to the j-th bin of histogram $hg(I)$.

$$hg_j(I) = \sum_v^n \beta_{uv} \text{ where } \beta_{uv} = \begin{cases} 1 & if\ s_v = b_u \\ 0 & otherwise \end{cases} \qquad (4)$$

The distance between histogram X and Y can be considered as the problem of finding the minimum difference of pair assignments between the two sets. The key issue is to determine the best one-to-one assignment between two sets such that the sum of all differences between two individual elements in a pair is minimized. Supposing there are n elements in histogram X and Y respectively, minimum difference of pair assignments of X and Y is defined as follows.

$$\Delta(X, Y) = \min_{X,Y}\left(\sum_{i,j=0}^{n-1} d(x_i, y_i)\right) \qquad (5)$$

where $x_i \in X$, $y_i \in Y$ and $d(\cdot)$ denotes the value of difference between two elements.

$$d(x_i, y_j) = \begin{cases} 0 & x_i, y_j \ are\ belonged\ to\ a\ same\ bin \\ 1 & otherwise \end{cases} \qquad (6)$$

There is little possibility that the number of key points of two images are equal to each other in practice. Therefore, we adopt least common multiple(LCM) to make the number of elements in the two histogram equal. In Eq.7, N_p and N_q are represented as the number of SIFT descriptors respectively.

$$hg_k^*(I_p) = \frac{LCM(N_p, N_q)}{N_p} \cdot hg_k(I_p) \qquad (7)$$

Hence, for a pair of images I_p and I_q, the local feature based image similarity can be computed by histogram sets of the two images as follows.

$$Sim_{Local}(I_p, I_q) = \frac{1}{L} \cdot \sum_{l=1}^{L} \frac{\Delta(hg^*(I_p)^l, hg^*(I_q)^l)}{LCM(N_p, N_q)} \qquad (8)$$

We submit tag t_i to Flickr, and then a photo collection $\delta(t_i)$ is obtained. Based on the photos returned from Flickr, The visual similarity between t_i and t_j is calculated using Eq.9 as follows.

$$\gamma_v(t_i, t_j) = \frac{\sum_{I_p \in \delta(t_i), I_q \in \delta(t_j)} Sim_{Local}(I_p, I_q)}{|\delta(t_i)| \cdot |\delta(t_j)|} \qquad (9)$$

To explore the complementary nature of semantic similarity and visual similarity, we combine them together.

$$\gamma(t_i, t_j) = \lambda \cdot \gamma_s(t_i, t_j) + (1-\lambda) \cdot \gamma_v(t_i, t_j) \qquad (10)$$

Tag visual similarity measuring method is shown in Fig.4.

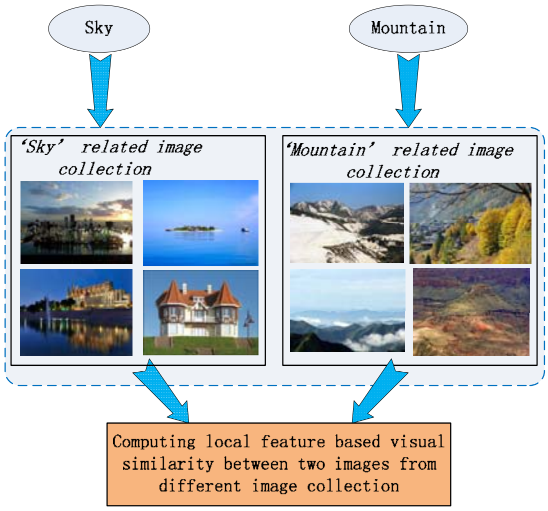

Sky
Mountain
'Sky' related image collection
'Mountain' related image collection
Computing local feature based visual similarity between two images from different image collection

Figure 4. Illustration of measuring visual similarity between tags.

B. Similarity between Tag and Image Region

We use normalized cuts[11] to segment images. After segmenting, an unlabeled photo can be represented by a region set Θ, the initial value of which is denoted as $\Theta^{(0)}$.

$$\Theta^{(0)} = \bigcup_{j=1}^{n} S_j,\ S_j \bigcap S_k = \varnothing \quad j \neq k \qquad (11)$$

where S_i is the descriptor of i-th segment.

To reduce computation cost and increase robustness against segmentation errors, we set a threshold β to combine visual similar segments into one region set. For instance, there are two similar segments in the image of Fig.2 which can represent 'sky'. We combine the segment S_x and S_y into region set R_z, if the following condition is satisfied.

$$\exp(-\frac{\|F(S_x) - F(S_y)\|^2}{\sigma^2}) < \beta \qquad (12)$$

where F_x and F_y are the visual feature vectors of region S_x and S_y. As is shown in Table 2, we totally extracted 168-dimension color and texture features as the low-level visual representation of the images.

TABLE II. THE 168-DIMENSION FEATURES WE USED

Feature category	Feature Name	Dimensions
Color	Color Correlogram	44
	Color Texture Moment	14
	Color Moment	6
Texture	Wavelet Features [21]	104

Next, we compute the visual feature vectors of R_z by a weighted fusion approach based on the area percentage scheme.

$$F(R_z) = \frac{\alpha_x}{\alpha_x + \alpha_y} F(S_x) + \frac{\alpha_y}{\alpha_x + \alpha_y} F(S_y) \qquad (13)$$

where α_i is the percentage of the image covered by segment S_i. Then, we update the current state of segment set $\Theta^{(t)}$ as follows.

$$\Theta^{(t+1)} = (\Theta^{(t)} - S_x - S_y) \bigcup R_z \qquad (14)$$

When the updating process of Θ converges, if there are some isolated segments in Θ, we transform the isolated segment descriptor S_k to R_k. Then, the converged state of Θ is denoted as $\Theta^{(f)}$, and $\Theta^{(f)} = \{R_1, R_2, \ldots, R_q\}$.

We measure the similarity between tags and the regions in Flickr photo by image content analyzing. Firstly, the initial tag t_i is submitted to Flickr, and then top M photos(denoted as ψ_{t_i}) returned from Flickr are obtained. Secondly, the photos belonged to ψ_{t_i} are segmented by normalized cuts. Finally, K-means is applied to cluster the segments of ψ_{t_i} into k parts $C^{t_i} = \{C_1^{t_i}, C_2^{t_i}, \ldots\ldots, C_k^{t_i}\}$ according to feature vector, and the centroid set are $M^{t_i} = \{m_1^{t_i}, m_2^{t_i}, \ldots\ldots, m_k^{t_i}\}$ where $m_l^{t_i}$ is the centroid of the l-th cluster in C^{t_i}. Therefore, the relevance between tag t_i and image region R_j is computed as follows.

$$\zeta(t_i, R_j) = \min\{\exp(-\frac{\|m_l^{t_i} - F(R_j)\|^2}{\sigma^2}), 1 \le l \le k\} \qquad (15)$$

C. Manifold-ranking Process

The manifold-ranking algorithm is a semi-supervised learning algorithm which explores the relationship among all the data points[12][13]. It has two versions for different tasks: to rank data points and to predict the labels of unlabeled data points. For the ranking task, it can be formulated as: given a set of points $\chi=\{x_1,\ldots,x_q,x_{q+1},\ldots,x_n\}\subset\mathbb{R}^m$, the first q points are the queries which form the query set, the remaining points are to be ranked according to their relevance to the queries. In this paper, the regions in $\Theta^{(f)}$ act as queries in the manifold-ranking process and the initial tags serve as the rest points in χ.

$$d(x_i,x_j)=\begin{cases}\gamma(x_i,x_j) & x_i,x_j\in\Gamma\\ \zeta(x_i,x_j) & x_i\in\Gamma,x_j\in\Theta^{(f)}\\ \exp(-\dfrac{\left\|F_{x_i}-F_{x_j}\right\|^2}{\sigma^2}) & x_i,x_j\in\Theta^{(f)}\end{cases} \qquad (16)$$

Let $d:\chi\times\chi\rightarrow\mathbb{R}$ denote a metric on χ which assigns each pair of points x_i and x_j a distance $d(x_i,x_j)$, and $f:\chi\rightarrow\mathbb{R}$ denote a ranking function which assigns to each point x_i a ranking score f_i. Finally, we define a vector $y=[y_1,\ldots,y_n]^T$ corresponding to the query set, in which $y_i=1$, if x_i is a query, and $y_i=0$ otherwise. The procedure of ranking the data points in [13] can be given as follows.

Manifold-Ranking Algorithm

(1) Sort the pair-wise distances among points in ascending order. Repeat connecting the two points with an edge according to the order until a connected graph is obtained.

(2) Form the affinity matrix W defined by $W_{ij}=\exp[-d^2(x_i,x_j)/2\sigma^2]$ if there is an edge linking x_i and x_j. Let $W_{ii}=0$.

(3) Symmetrically normalize W by $S=D^{-1/2}WD^{-1/2}$ in which D is the diagonal matrix with (*i*, *i*)-element equal to the sum of the *i*-th row of W.

(4) Iterate $f(t+1)=\alpha Sf(t)+(1-\alpha)y$ until convergence, where α is a parameter in [0, 1) and $f(0)=y$.

(5) Let f^* denote the limit of the sequence $\{f(t)\}$. Rank each point x_i according to its ranking score f_i^*.

After manifold-ranking process, top ranked tags set(denoted as Γ^T) are reserved.

TABLE III. EXAMPLES OF RELATED TAGS GENERATED BY FLICKR API

Tag	Related tags generated by Flickr API
Sky	clouds blue sunset water tree sea sand sun nature night cloud trees building light landscape cielo architecture white beach moon red silhouette sunrise grass green canon ocean nikon orange flower hdr boat lake color field winter yellow evening azul dusk snow rain mountain road city pink flowers nubes
Island	sea beach water sky blue clouds ocean sunset sand sun nature landscape boat summer travel vacation trees italy coast rocks greece green waves mar cloud mare holiday paradise tree canon nikon boats thailand bay orange light tropical italia reflection white ship playa

IV. EXPANDING THE TOP RANKED INITIAL TAGS

In this section, we will demonstrate how to expand the top ranked user-supplied tags generated from section 4 (shown in Fig.5). We apply a Flickr API to obtain related tags, and then use a weighted voting policy to expand the top ranked tags. The main idea of our tags expanding policy lies in that the tags which are more relevant to the high ranked initial tags would be more suitable.

A. Obtaining Related Tags by Flickr API

According to the Flickr's Related Tag API which is named flickr.tags.getRelated, each tag has a list of "related" tags. Table 3 shows some tags and their related tags by Flickr API.

B. Expanding Tags by Weighted Voting

Supposing tag t_i is denoted as the *i*-th tag in the top ranked tag set Γ^T, the related tags of which are represented as $R(t_i)$. We merge all the related tags together and eliminate duplicated tags to build up the candidate expanding tag set Γ^E.

$$\Gamma^E=\bigcup\nolimits_{t_i\in\Gamma^T}R(t_i)=\{e_1,e_2,\ldots,e_k\} \qquad (17)$$

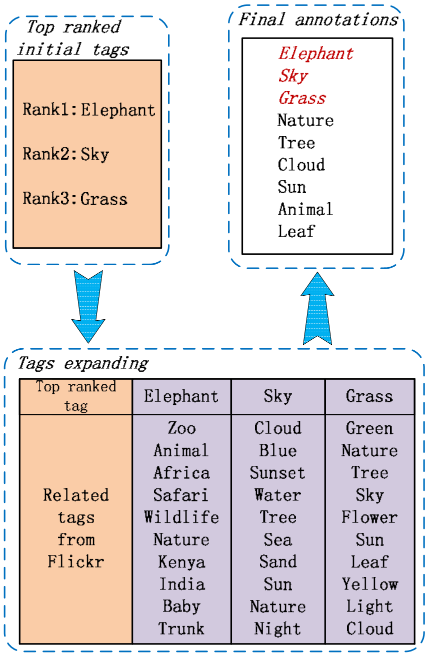


Figure 5. Framework of tag expanding method.

To make the expanding tags more relevant to the photo to be annotated, two factors are considered in our weighted voting policy. Firstly, the influence of higher ranked initial tags to voting results is boosted. Secondly, the semantic relevance between tags is taken into account as well. The weighted voting strategy computing a score for candidate tag e_j is designed as follows.

$$wv(e_j, t_i) = \begin{cases} \dfrac{\left|\Gamma^T\right| - i + 1}{\left|\Gamma^T\right|}, & e_j \in R(t_i) \\ 0 \;, & e_j \notin R(t_i) \end{cases} \qquad (18)$$

Based on voting score calculated from Eq.18, the final ranking score of each candidate expanding tag is computed as follows.

$$score(e_j) = \sum\nolimits_{t_i \in \Gamma^T} wv(e_j, t_i) \cdot \gamma(e_j, t_i) \qquad (19)$$

According to the score of related tags calculated by Eq.19, the tags with high scores would be reserved as expanding tags. Finally, the final annotations are constructed by combining top ranked user-supplied tags with expanding tags.

V. Experimental Results and Analysis

We collected 2000 Flickr photos as experimental dataset, with 20 categories and 100 photos in each category. Each photo category is built up by submitting a popular tag of Flickr and we download 100 photos which have at least 10 initial tags. The reason we select the photos which is relevant to the query lies in that the photos' visual content are of diversity. Flickr provides a service to give users the relevant photos to users' query. 20 popular photo categories are used in this experiment, including beach, city, flower, park, sky, snow, street, summer, sunset, travel, water, wedding, winter, tree, island, bird, mountain, dog, cat and automobile. To avoid experiment results being subjective, we arrange 20 volunteers to judge the performance of our approach, and then integrate all volunteers' opinions to get the overall performance evaluation. We design three different experiment schemes to test the performance of our approach. There are four parameters to be set in the algorithm: λ , α , σ and the number of iteration. In this experiment, the above four parameters are set as 0.4, 0.99, 1 and 50 respectively.

In experiment 1, we evaluate the performance of the proposed tag ranking method. Normalized Discounted Cumulative Gain (NDCG) is used as the metric to measure ranking performance[14]. NDCG metric is a retrieval measurement devised specifically for Web search evaluation. It is based on human judgments. Human judges rate on how relevant each retrieval result is on an *n*-point scale. For a given query *q*, the NDCG is computed as shown in Eq.20.

$$NDCG(q) = M_q \sum_{j=1}^{K} \frac{2^{r(j)} - 1}{\log(1 + j)} \qquad (20)$$

where M_q is a normalization constant calculated so that the perfect ordering would obtain NDCG of 1, and $r(j)$ is an integer representing the relevancy rated by human. NDCG rewards relevant elements in the top ranked results more heavily than those ranked lower and punishes irrelevant elements by reducing their contributions to NDCG. In this experiment, each annotation of an image is labeled as one of the five levels: Strong Relevant (score 5), Relevant (score 4), Partially Relevant (score 3), Weakly Relevant (score 2), and Irrelevant (score 1).

To compute NDCG, 20 volunteers are assigned to rate the relevancy of each annotation with the score from 1 to 5. After computing the NDCG value of each image's annotation list, we can average them to obtain an overall performance evaluation of the annotation ranking capability. The NDCG performance under different photo categories is shown in Table 4. We select the top 10 tags in all cases, which include 1) Initial tags, 2) Liu's method[1] and 3) our approach. As initial tags were not ranked, we use the user submitting order as the tag ranking results.

TABLE IV. NDCG PERFORMANCE UNDER DIFFERENT PHOTO CATEGORIES

	Beach	**City**	**Flower**	**Park**	**Sky**	**Snow**	**Street**	**Summer**	**Sunset**	**Travel**
Initial tags	0.586	0.574	0.588	0.595	0.515	0.525	0.599	0.536	0.488	0.512
Liu's method	0.821	0.730	0.812	**0.754**	0.725	0.757	0.726	**0.674**	0.621	**0.595**
Our approach	**0.858**	**0.734**	**0.878**	0.750	**0.733**	**0.787**	**0.733**	0.629	**0.643**	0.594
	Water	**Wedding**	**Winter**	**Tree**	**Island**	**Bird**	**Mountain**	**Dog**	**Cat**	**Automobile**
Initial tags	0.499	0.601	0.571	0.633	0.547	0.636	0.542	0.525	0.560	0.531
Liu's method	0.540	0.815	0.574	0.785	0.774	0.835	0.745	0.812	0.796	0.799
Our approach	**0.574**	**0.872**	**0.581**	**0.834**	**0.792**	**0.872**	**0.824**	**0.911**	**0.877**	**0.905**

From Table 4, we can see that the tag ranking performance of our approach is superior to initial tags. Compared with Liu's method, our approach performs better in some situations, such as wedding, bird, dog, cat, automobile etc. We can find that in these cases there are some salient and visual similar objects in the photos. The reason lies in two aspects. Firstly, using local features, our approach could effectively recognize salient objects. Secondly, our tag similarity measuring policy is more reasonable, as we use NFD to represent tag concurrence similarity and adopt SIFT descriptors based method to measure image visual similarity.

Experiment 2 shows the tag expanding performance of our weighted voting scheme. We choose expanding tags from candidate expanding tags by a ranking score computing (shown in Eq.19). Hence, NDCG can also be used in tag expanding performance measuring. Table 5 demonstrates tag expanding performance, which compares tag expanding policy without weighted voting. To abandon the weighted voting, we should modify the first case of Eq.19. If $e_j \in R(t_i)$, we let $wv(e_j, t_i) = 1$.

From Table 5, some conclusions can be drawn. 1) Tag expanding performance highly depends on the ranking accuracy of initial tags. 2) Our weighted voting policy can enhance tag expanding performance evidently.

TABLE V. NDCG PERFORMANCE FOR DIFFERENT TAG EXPANDING POLICY

	Beach	City	Flower	Park	Sky	Snow	Street	Summer	Sunset	Travel
Initial tags after ranking	0.858	0.734	0.878	0.750	0.733	0.787	0.733	0.629	0.643	0.594
Without Voting	0.870	0.757	0.857	0.792	0.753	0.838	0.764	0.680	0.653	0.603
Our approach	0.956	0.820	0.895	0.838	0.779	0.835	0.829	0.679	0.715	0.651
	Water	**Wedding**	**Winter**	**Tree**	**Island**	**Bird**	**Mountain**	**Dog**	**Cat**	**Automobile**
Initial tags after ranking	0.574	0.872	0.581	0.834	0.792	0.872	0.824	0.911	0.877	0.905
Without Voting	0.585	0.895	0.607	0.848	0.807	0.909	0.843	0.890	0.859	0.911
Our approach	0.639	0.928	0.644	0.876	0.860	0.906	0.890	0.950	0.949	0.953

Experiment 3 test the performance of our photo annotating approach and the effectiveness of the two-level tag similarity measuring mechanism. Two metrics are adopted in this experiment. The first metric is average precision of top *N* annotations (AP@N) in extending annotations, which evaluates how many annotations in top *N* position are relevant to the unlabeled image. To compute AP@N, a boolean type function is defined as follows.

$$ifTrue(\alpha_i) = \begin{cases} 1, & \alpha_i \text{ is correct} \\ 0, & otherwise \end{cases} \qquad (21)$$

Suppose the number of volunteers is *U*. AP@N is computed in Eq.22.

$$AP@N = \frac{1}{U}\sum_{j=1}^{U}\frac{\sum_{i=1}^{N} ifTrue(\alpha_i)}{N} \qquad (22)$$

Another metric is the average coverage rate of top *N* annotations (AC@N) in extending annotations, which estimates if top *N* annotations at least include one relevant annotation. A boolean function is also defined in advance:

$$ifCover(N) = \begin{cases} 1, & at\ least\ one\ revelant\ annotation\ in\ Top\ N\ annotations \\ 0, & otherwise \end{cases}$$

(23)

Averaging all users' opinions, AC@N is solved as is shown in Eq.24.

$$AC@N = \frac{1}{U}\sum_{j=1}^{U} ifCover(N) \qquad (24)$$

We adopt two different tag similarity measuring method in our approach to test the performance of the two-level tag similarity measuring method. Two methods are compared to our approach which includes: 1) only using concurrence similarity to compute tag similarity(denoted as CS_only), 2) only using visual similarity to compute tag similarity(denoted as VS_only). Fig.6 and Fig.7 illustrate average annotation precision and average annotation coverage of all the Flickr photos we have downloaded. In this experiment, the final annotations include three top ranked initial tags which located in the first three positions, and the other positions in final annotations are made up of expanding tags.

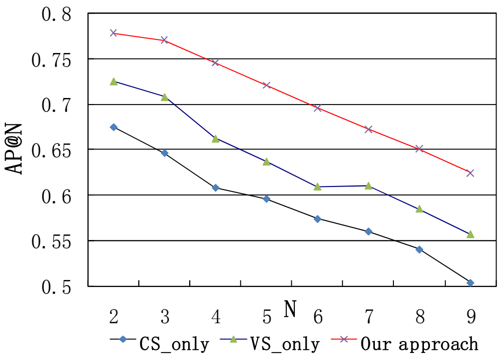


Figure 6. AP@N performance under different settings.

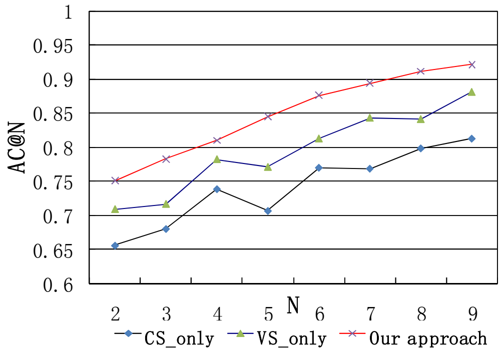


Figure 7. AC@N performance under different settings.

From the experimental results shown in Fig.6 and Fig.7, several conclusions can be drawn: 1) Adopting the two-level tag similarity measuring method could enhance the overall annotation performance in both AP@N and AC@N. 2) After tag ranking and expanding process, more relevant annotations are located in top positions.

VI. CONCLUSIONS AND FUTURE WORKS

Annotating photos through user-supplied tags mining is a popular way to index and organize photos. This paper proposes a Flickr photo-oriented automatic image annotation approach by manifold-ranking based tag ranking and weighted voting based tag expanding. We perform a manifold-ranking based method to rank initial tags which are obtained by pruning noisy tags from user-supplied tags. Next, using the relevant tags which Flickr API provides, tags are expanded by computing relevance score. Finally, combining top ranked initial tags and expanding tags, we can obtain the final annotations. We design three experiment schemes to test the proposed approach.

In the future, we would like to extend our work in the following directions. 1) We will try other methods to measure image visual similarity and semantic similarity between words. 2) Related information of the users who upload the photos, such as user interests or the groups of which users are belonged to, will be used as metadata to enhance annotation performance. 3) We will extend the scale of testing dataset in experiments.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No.60970048), Research Foundation of Shandong Economic University ("Research on Key Problems of Automatic Image Annotation", 2008), and Project of Shandong Province Higher Educational Science and Technology Program (Grant No.J10LG69).

REFERENCES

[1] D. Liu, X. S. Hua, L. Yang, et al. "Tag ranking," In Proceedings of the 18th international conference on World Wide Web, pp.351-360, 2009.

[2] Börkur Sigurbjörnsson and Roelof van Zwol. "Flickr tag recommendation based on collective knowledge," In Proc. ACM WWW 2008, pp.327-336.

[3] H. Xu, X. Zhou, M Wang, et al. "Exploring Flickr's related tags for semantic annotation of web images," Proceeding of the ACM CIVR, 2009.

[4] L. Wu, X.-S. Hua, N. Yu, et al. "Flickr distance," ACM MM, pp.31-40, 2008.

[5] P. Schmitz. "Inducing ontology from flickr tags," WWW 2006.

[6] M. Chen, M. H. Chang, P. C. Chang, et al. "SheepDog-Group and Tag Recommendation for Flickr Photos by Automatic Search-based Learning," In Proceeding of ACM MM, 2008.

[7] L. Wu, L. J. Yang, N. H. Yu and X. S. Hua. "Learning to Tag," In Proceeding of ACM WWW 2009.

[8] M. Ames and M. Naaman. "Why We Tag: Motivations for Annotation in Mobile and Online Media," In Proceeding of ACM SIGCHI 2007.

[9] L. S. Kennedy, S. F. Chang, and I. V. Kozintsev. "To Search or To Label? Precdicting the Performance of Search-Based Automatic Image Classifiers," In Proceedings of ACM MIR 2006.

[10] Duygulu, P., Barnard, K., de Freitas, J., Forsyth, D.A. "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," Proc. ECCV 2002, pp.97-112.

[11] J. Shi and J. Malik. "Normalized cuts and image segmentation," In Proc. CVPR 1997, pp.731-743.

[12] D. Zhou, O. Bousquet, T. N. Lal, J. Weston and B. SchÖlkopf. "Learning with local and global consistency," In Proceedings of NIPS 2003.

[13] D. Zhou, J. Weston, A. Gretton, O. Bousquet and B. SchÖlkopf. "Ranking on data manifolds," In Proceedings of NIPS 2003.

[14] K. Jarvelin, and J. Kekalainen. "IR evaluation methods for retrieving highly relevant documents," In Proc. ACM SIGIR 2000.

[15] Rudi L. Cilibrasi, Paul M.B. Vitányi. "The Google Similarity Distance," *IEEE Trans. on Knowledge and Data Engineering*. 19(3), pp.370-383, 2007.

[16] D. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, 60(2), pp.91-110, 2004.

[17] Linde, Y., Buzo, A., Gray, R. "An Algorithm for Vector Quantizer Design," *IEEE Trans. on Communications*, 28, pp.84-94, 1980.

[18] P. Indyk, R. Motwani, P. Raghavan, and S. Vempala, "Approximate Nearest Neighbor: Towards Removing the Curse of Dimensionality," Proc. 30th ACM Symp. Computational Theory, pp.604-613, 1998.

[19] M. Datar, N. Immorlica, P. Indyk, and V.S. Mirrokni, "Locality-Sensitive Hashing Scheme Based on p-Stable Distributions," Proc.20th Symp. Computational Geometry, pp.253-262, 2004.

[20] Sung-Hyuk Cha, Sargur N. Srihari, "On measuring the distance between histograms," *Pattern Recognition*, Volume 35, Issue 6, June 2002, pp.1355-1370.

[21] Chang, T., and Kuo, C.C.J. "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. on Image Processing*, vol. 2, no. 4, pp.429-441, 1993.

Zheng Liu, born in 1980, earned a B.S. and M.S. degree in Computer Science & Technology from Shandong University, in 2002 and 2005 respectively. After graduate school, he joined school of Computer Science & Technology, Shandong Economic University in 2005. Now he is pursuing his Ph.D. degree in Shandong University, and works with the Information Retrieval Group. His current research interests include machine learning, pattern recognition and multimedia data mining.

Hua Yan, born in 1973, received the Ph.D. degree in communication and information system from Shandong University, Jinan, China, in 2007. Now she is an associate professor in Shandong Economic University. Her current research interests include various aspects of image/video processing and its application.