Bridging the Semantic Gap Between Image Contents and Tags

Hao Ma, Jianke Zhu, *Member, IEEE*, Michael Rung-Tsong Lyu, *Fellow, IEEE*, and Irwin King, *Senior Member, IEEE*

***Abstract*—With the exponential growth of Web 2.0 applications, tags have been used extensively to describe the image contents on the Web. Due to the noisy and sparse nature in the human generated tags, how to understand and utilize these tags for image retrieval tasks has become an emerging research direction. As the low-level visual features can provide fruitful information, they are employed to improve the image retrieval results. However, it is challenging to bridge the semantic gap between image contents and tags. To attack this critical problem, we propose a unified framework in this paper which stems from a two-level data fusions between the image contents and tags: 1) A unified graph is built to fuse the visual feature-based image similarity graph with the image-tag bipartite graph; 2) A novel random walk model is then proposed, which utilizes a fusion parameter to balance the influences between the image contents and tags. Furthermore, the presented framework not only can naturally incorporate the pseudo relevance feedback process, but also it can be directly applied to applications such as content-based image retrieval, text-based image retrieval, and image annotation. Experimental analysis on a large Flickr dataset shows the effectiveness and efficiency of our proposed framework.**

***Index Terms*—Content-based image retrieval, image annotation, random walk, text-based image retrieval.**

I. Introduction

IMAGE retrieval has been adopted in most of the major search engines, including Google, Yahoo!, Bing, etc. A large number of image search engines mainly employ the surrounding texts around the images and the image names to index the images. However, this limits the capability of the search engines in retrieving the semantically related images using a given query. On the other hand, although the current state-of-the-art in content-based image retrieval is progressing, it has not yet succeeded in bridging the semantic gap between human concepts, e.g., keyword-based queries, and low-level visual features that are extracted from the images [22], [36]. Hence, it has become an urgent need for developing novel and effective paradigms that go beyond these conventional approaches or retrieval models.

Manuscript received October 31, 2009; revised February 28, 2010; accepted April 25, 2010. Date of publication May 27, 2010; date of current version July 16, 2010. This work was supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 4128/08E and CUHK 4154/09E). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Abdulmotaleb El Saddik.

The authors are with the Department of Computer Science and Engineering, Chinese University of Hong Kong, Kowloon, Hong Kong (e-mail: hma@cse.cuhk.edu.hk; jkzhu@cse.cuhk.edu.hk; lyu@cse.cuhk.edu.hk; king@cse.cuhk.edu.hk.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2010.2051360

Tags:
Hong Kong
night
IFC
Bank of China Tower
skyline
travel

Fig. 1. Example image with its tags.

Recently, with the prevalence of Web 2.0 applications and social games, more and more users contribute numerous tags to Web images, such as Flickr [17], ESP games [14], etc. These tags provide the meaningful descriptors of images, which are especially important for those images containing little or no textual context. The success of Flickr proves that users are willing to provide this semantic context through manual annotations. Recent user studies on this topic reveal that users do annotate their photos with the motivation to make them better accessible to the general public [1]. Fig. 1 shows an example image extracted from Flickr with its user-generated tags. This is a photo taken in Hong Kong, and is described by users with the tags *Hong Kong*, *night*, *IFC* (a building name, standing for International Finance Centre), *Bank of China Tower*, and *skyline*, which are all semantically relevant to this image. However, it is very difficult for the current content-based image retrieval methods to produce such meaningful results. Hence, the tag data is an ideal source for improving many tasks in image retrieval.

Unfortunately, tags inevitably contain the injected noise in the manual labeling process. As shown in Fig. 1, the last tag associated with this photo is *travel*. To the owner who submitted this photo, obviously, this tag is not a noisy tag since this photo probably was taken when the owner was traveling to the city. But in terms of the image retrieval tasks, most probably the tag *travel* is a noise, since it is a too general term. Other popular tags in Flickr like *nature*, *2008* also belong to this category. Therefore, simply using tags in image retrieval tasks is not a reliable and reasonable solution; the visual information of images should also be taken into consideration to improve the image search engines since visual information gives the most direct correlations between images, and 80% of human cognition comes from visual information [40].

In this paper, we investigate the research problem on how to incorporate both image content and tag information into image retrieval and annotation tasks. To take advantages of both the visual information and user-contributed tags for image retrieval, we need to tackle two main challenges. The first challenge is how to bridge the semantic gap between image contents and image tags. Essentially, visual features and tags are two different but closely related aspects of images. Although content-based image retrieval using visual features has been extensively studied since the 1990s, the semantic gap between low-level image features and high-level semantic concepts is still the key hindrance towards the effectiveness of content-based image retrieval systems [39]. The second challenge is how to create a scalable and effective algorithm. There are a huge amount of images on the Web, and more and more new photos are uploaded to photo sharing Web sites like Flickr everyday by numerous independent Web users. Hence, a scalable and effective algorithm is necessary to analyze both the visual information and the tags of images.

Aiming at the above challenges, we propose a unified framework for performing tasks related to image retrieval, including content-based image retrieval, text-based image retrieval, and image annotation. This framework relies on a two-level data fusion between image contents and tags. Based on the global features extracted from every image, we first infer an image similarity graph, and then form a hybrid graph with the image-tag bipartite graph. In this hybrid graph, one part of the weighted edges are connecting different images and the weights represent the similarities between them, while the other part of the weighted edges are bonding the images and tags with the weights reflecting the co-occurrent frequencies. After building the hybrid graph, we then propose a novel and effective random walk model that employs a fusion parameter to balance the importance between the image contents and the tags. The fusion parameter determines whether to accelerate the diffusion of random walks of image-tag subgraph, or to accelerate the walks of image-image subgraph. Moreover, our framework also provides a natural solution for including the pseudo relevant feedback into image retrieval and annotation tasks. The experimental results of three applications on a large Flickr dataset show the advantage of our proposed framework.

The rest of the paper is organized as follows. We review related work in Section II. Section III describes the proposed unified framework, including the global feature extraction, the hybrid graph construction, and the random walk model. In Section IV, we demonstrate the empirical analysis of our framework on three image retrieval applications. Finally, conclusions and future work are given in Section V.

II. Related Work

Considerable research efforts [5], [11], [12], [19], [26], [36] have been devoted to address attacking the semantic gap between low-level features and high-level semantic concepts, which is the key hindrance in content-based image retrieval [10], [35].

Machine learning techniques have been shown as one way to bridge the semantic gap between the image features and semantic concepts. The recent research literatures have been significantly interested in employing graphical models and distance metric learning algorithms. The work in [5] is inspired from the natural language processing methods, in which the process of building the relation between the visual features and the keywords is analogous to a language translation. Similarly, Djeraba [11] tries to learn the associations between the visual features and the semantic descriptions via a visual dictionary. As for the distance metric learning, it is mainly employed to construct the semantic map [39] which learns a distance measure to approximate the similarity in the textual space. Therefore, the learnt similarity measure can be further employed in the image annotation task. Moreover, the semantic concept relationship can be captured by the visual correlation between concepts [40], [34], which is essential to the concept clustering, semantic distance estimation, and image annotation. Additionally, learning with relevance feedbacks, which takes advantage of the users' interaction to improve the retrieval performance, has been extensively studied [35], [38]. One disadvantage for the learning-based methods is its generalization capability. A remedy is to raise the total number of representative training examples; however, this requires more manually labeled data and increases the computational cost significantly.

Another approach to the semantic gap issue is to take advantage of the advance in computer vision domain, which is closely related to object recognition and image analysis. Duygulu *et al.* [12] present a machine translation model which maps the keyword annotation onto the discrete vocabulary of clustered image segmentations. Moreover, Blei and Jordan [6] extend this approach through employing a mixture of latent factors to generate keywords and blob features. Jeon *et al.* [21] reformulate the problem as cross-lingual information retrieval, and propose a cross-media relevance model to the image annotation task. In contrast to the image-based and region-based methods, the image content is represented by the salient objects in [16], which can achieve automatic image annotation at the content level. A hierarchical classification frame is proposed in [15], which employed salient objects to characterize the intermediate image semantics. Those salient objects are defined as the connected image regions that capture the dominant visual properties linked to the corresponding physical objects in an image. However, these methods rely on the results of image segmentation and salience detection, which are sensitive to the illumination conditions and cluttered background. In most recent, bag-of-words representation [8], [42] of the local feature descriptors demonstrated promising performance in calculating the image similarity. To deal with the high-dimensionality of the vector feature space, the efficient hashing index methods have been investigated in [8] and [24]. These approaches did not take consideration of the tag information which is very important for the image retrieval task.

Apart from its connection with research work in content-based image retrieval, our work is also related to the broad research topic in graph-based methods. Graph-based methods are intensively studied with the aim of reducing the gap of the visual features and semantic concept. In [23], the images are represented by the attributed relational graphs, in which each node in the graph represents an image region and each edge represents a relation between two regions. In [18], an image is represented as a sequence of feature-vectors characterizing low-level visual features, and is modeled as if it was stochastically generated by a hidden Markov model, whose states represent concepts. Most recently, Jing and Baluja [22] present an intuitive graph model-based method for product image search. They directly view images as documents and their similarities as probabilistic visual link. Moreover, the likelihood of images is estimated by a random walk algorithm on the image similarity graph. The image similarity is based on the local feature matching using SIFT descriptor [27]; unfortunately, this incurs heavy computational cost. Recently, several random walk-based methods have been proposed in image and video retrieval tasks [4], [9], [20]. However, the image-tag [9] and video-view graphs [4] based approaches did not take consideration of the contents of images or videos, which lose the opportunity to retrieve more accurate results. In [20], a re-ranking scheme is developed using random walk over the video story graph. Multiple-instance learning can also take advantage of the graph-based representation [37] in the image annotation task.

Our work is also related to recommender systems since it aims to recommend relevant images and tags. However, different with traditional recommendation or collaborative filtering algorithms [28]–[30], our work does not have user-item rating information. Hence, in some sense, the problem we study in this work is more difficult than some of the traditional recommendation problems.

Instead of relying on the complicated models and representative training examples in the machine learning-based methods, we propose an effective and efficient framework based on Markov random walk [33], which can take advantage of both image visual contents and image tags. Our method does not need to train any functions or models, and can be easily scaled to very large dataset.

III. Unified Framework

In this section, we detail our framework, including how to extract global features, how to build the hybrid graph based on visual features and tags, and how to perform a random walk on it.

A. Global Feature Extraction

The global feature representation techniques have been extensively studied in image processing and content-based image retrieval. In contrast to the local feature-based approaches [22], the global feature is very efficient in computation and storage due to its compact representation. A wide variety of global feature extraction techniques have been proposed in the past decade. In this paper, we extract four kinds of effective global features.

- **Grid color moment**. We adopt the grid color moment to extract color features from images. Specifically, an image is partitioned into a 3×3 grids. For each grid, we extract three kinds of color moments: color mean, color variance, and color skewness in each color channel (R, G, and B), respectively. Thus, an 81-dimensional grid color moment vector is adopted for color features.
- **Local binary pattern (LBP)**. The local binary pattern [31] is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. In our experiment, a 59-dimensional LBP histogram vector is adopted.
- **Gabor wavelets texture**. To extract Gabor texture features, each image is first scaled to 64×64 pixels. The Gabor wavelet transform [25], [43] is then applied on the scaled image with five levels and eight orientations, which results in 40 subimages. For each subimage, three moments are calculated: mean, variance, and skewness. Thus, a 120-dimensional vector is used for Gabor texture features.
- **Edge**. An edge orientation histogram is extracted for each image. We first convert each image into a grayscale image, and then employ a Canny edge detector [7] to obtain the edge map for computing the edge orientation histogram. The edge orientation histogram is quantized into 36 bins of 10 degrees each. An additional bin is used to count the number of pixels without edge information. Hence, a 37-dimensional vector is used for shape features.

In total, a 297-dimensional vector is used to represent all the global features for each image in the dataset, which is further normalized to zero mean and unit variance. Note that this feature representation has shown the promising performance on the duplicate image retrieval task [44], [45] which requires the accurate similarity measure for the image pairs.

B. Hybrid Graph Construction

Once the visual features are extracted, we can build the image similarity graph. Let v represent the dimensionality of each image, and let $\mathcal{D}$ denote the total image set. For each image $d_p \in \mathcal{D}, p \in [1, |\mathcal{D}|]$, let the v-dimentional vector $\mathbf{d}_p$ represent the image feature vector corresponding to image d_p. We employ the cosine function to measure the similarity between two images d_p and d_q:

$$\mathrm{Sim}(d_p, d_q) = \frac{\mathbf{d}_p \cdot \mathbf{d}_q}{||\mathbf{d}_p|| ||\mathbf{d}_q||}. \tag{1}$$

We then build the image similarity graph based on the calculated similarities. Usually, there are several methods to construct similarity graphs, including kNN graph, ϵNN graph, exp-weighted graph, etc. As reported in [46], kNN graph tends to perform well empirically. Hence, for an image d_p, we employ the k most similar images as its neighbors. More specifically, if an image d_q is in the k-nearest-neighborhood of image d_p, then we create a directed edge from node d_p to node d_q, and the weight is the similarity $\mathrm{Sim}(d_p, d_q)$. This kNN graph is an asymmetric graph since if node d_q is in the k-nearest-neighborhood of node d_p, it does not mean node d_p is also in the k-nearest-neighborhood of node d_q. Fig. 2(a) illustrates an example kNN graph with $k = 1$. Note that it is time-consuming to find the most similar images by brute-force searching in a very large dataset. Fortunately, we can take advantage of the nearest neighbor searching method proposed in [2] to efficiently build the image similarity graph.

Beside the image similarity graph, we also build the image-tag bipartite graph. However, we cannot simply incorporate the image-tag bipartite graph into our framework. This is because the bipartite graph is an undirected graph, which cannot accurately interpret the relationships between images and tags. To tackle this problem, we need to convert it into a directed graph. As shown in Fig. 2(b), the left part of the bipartite graph represents the image nodes, while the right part denotes the tag nodes. In the converted graph, every undirected edge in the original bipartite graph is converted into two directed edges. The weight on a new directed image-tag edge is normalized by

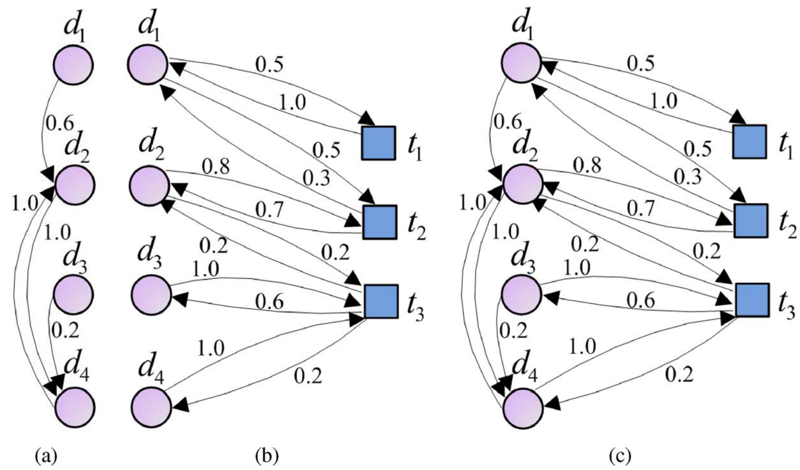


Fig. 2. Hybrid graph construction.

the total number of times that the image is tagged, while the weight on a directed tag-image edge is normalized by the total number of times that this tag has been assigned.

After building the image similarity graph and the image-tag directed graph, we consolidate these two graphs, and create a directed hybrid graph, as shown in Fig. 2(c). This directed hybrid graph forms the foundation of our random walk model that will be introduced in the next section.

C. Random Walk Model

Markov random walk model has been extensively studied in many Web applications. In this section, we introduce a novel random walk model on our hybrid graph that can smoothly employ the visual and the textual information into several image retrieval tasks.

Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ denote a directed hybrid graph, where $\mathcal{V} = \mathcal{D} \cup \mathcal{T}$ is the vertex set, and $\mathcal{D}$ represents the set of image nodes while $\mathcal{T}$ denotes the set of tag nodes. $\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^*$ is the edge set which consists of two types of edges. If the edge $\mathcal{E}_{ij}$ is in the edge set $\mathcal{E}^+$, then $i \in \mathcal{D}$ and $j \in \mathcal{D}$. If the edge $\mathcal{E}_{ij}$ is in the edge set $\mathcal{E}^*$, then $i \in \mathcal{D}, j \in \mathcal{T}$, or $i \in \mathcal{T}, j \in \mathcal{D}$.

For all the edges in the edge set $\mathcal{E}^*$, we define the transition probability $P''_{t+1|t}(j|i)$ from node $\mathcal{V}_i$ to node $\mathcal{V}_j$ as $C_{ij}/\sum_{p\in\mathcal{D}} C_{ip}$ or $C_{ij}/\sum_{p\in\mathcal{T}} C_{ip}$. If $i \in \mathcal{D}, j \in \mathcal{T}$, then C_{ij} is the number of times that the tag node $\mathcal{V}_j$ has been assigned to the image node $\mathcal{V}_i$, while $\sum_{p\in\mathcal{T}} C_{ip}$ is the total number of times that the image node $\mathcal{V}_i$ has been tagged. If $i \in \mathcal{T}, j \in \mathcal{D}$, then C_{ij} is the number of times that the tag node $\mathcal{V}_i$ has been assigned to the image node $\mathcal{V}_j$, while $\sum_{p\in\mathcal{D}} C_{ip}$ is the total number of times that the tag node $\mathcal{V}_i$ has been assigned to all the images. Actually, this consideration denoises the popular tags with little meanings like "nature" and "travel" since the weights on the edges starting from these nodes will be very small. The notation $P''_{t+1|t}(j|i)$ denotes the transition probability from node $\mathcal{V}_i$ at time step t to node $\mathcal{V}_j$ at time step $t+1$. While the counts C_{ij} are symmetric, the transition probabilities $P''_{t+1|t}(j|i)$ generally are not, because the normalization varies across different nodes.

For other edges in the edge set $\mathcal{E}^+$, we define the transition probability $P''_{t+1|t}(j|i)$ from node $\mathcal{V}_i$ to node $\mathcal{V}_j$ as $\mathrm{Sim}(\mathcal{V}_i, \mathcal{V}_j)/\sum_{p\neq i} \mathrm{Sim}(\mathcal{V}_i, \mathcal{V}_p)$, where $\mathrm{Sim}(\mathcal{V}_i, \mathcal{V}_j)$ is the image visual similarity between the image nodes $\mathcal{V}_i$ and $\mathcal{V}_j$ defined in (1). This is slightly different from the example we show in Fig. 2, since we normalize the similarities here.

In general, the transition probability is

$$P''_{t+1|t}(j|i) = \begin{cases} \dfrac{C_{ij}}{\sum_{p\in\mathcal{T}} C_{ip}}, & \mathcal{E}_{ij} \in \mathcal{E}^*, i \in \mathcal{D} \\ \dfrac{C_{ij}}{\sum_{p\in\mathcal{D}} C_{ip},} & \mathcal{E}_{ij} \in \mathcal{E}^*, i \in \mathcal{T} \\ \dfrac{\mathrm{Sim}(\mathcal{V}_i, \mathcal{V}_j)}{\sum_{p} \mathrm{Sim}(\mathcal{V}_i, \mathcal{V}_p),} & \mathcal{E}_{ij} \in \mathcal{E}^+. \end{cases} \tag{2}$$

The random walk can only diffuse through the links that connect nodes in a given graph; in fact, there are random relations among different nodes even if these nodes are not connected. For example, in the similarity relations on the image similarity graph, we explicitly calculate the similarities between images based on (1). Actually, there are some implicit hidden similarity relations among these images that cannot be observed or captured. Hence, to capture these relations, without any prior knowledge, we propose to add a uniform random relation among different nodes. More specifically, let γ denote the probability that such phenomena happen, and $(1-\gamma)$ is the probability of taking a "random jump". Without any prior knowledge, we set $\mathbf{g} = (1/n)\mathbf{1}$, where $\mathbf{g}$ is a uniform stochastic distribution vector, $\mathbf{1}$ is the vector of all ones, and n is the number of nodes. Based on the above consideration, we modify our model to use the following transition probability matrix:

$$\mathbf{P}' = \gamma\mathbf{P}'' + (1-\gamma)\mathbf{g}\mathbf{1}^T \tag{3}$$

where matrix $\mathbf{P}''$ is the transition probability matrix with the entry of the ith row and the jth column defined in (2). Following the setting of γ in PageRank [13], [32], we set $\gamma = 0.85$ in all of our experiments conducted in Section IV.

Our graph is a hybrid graph which consists of two totally different subgraphs: image similarity graph and image-tag bipartite graph. Intuitively, the contributions of these two graphs are not likely the same. This indicates that in some applications, the image-tag subgraph is more important than the image similarity subgraph, while in other applications, the image similarity subgraph should contribute more. Hence, in order to endow more flexibility to our random walk model, we employ a fusion parameter λ to the transition matrix $\mathbf{P}'$ introduced in (3). We define the transition probability matrix $\mathbf{P}$ with the entry $P_{t+1|t}(j|i)$ from node $\mathcal{V}_i$ to node $\mathcal{V}_j$ as

$$P_{t+1|t}(j|i) = \begin{cases} \lambda P'_{t+1|t}(j|i), & \mathcal{E}_{ij} \in \mathcal{E}^* \\ (1-\lambda)P'_{t+1|t}(j|i), & \mathcal{E}_{ij} \in \mathcal{E}^+. \end{cases} \quad (4)$$

The parameter λ plays as a very important role in our random walk model, which defines how fast the random walk diffuses on the two subgraphs. Following a physical intuition, when $\lambda = 1$, the random walk only performs on the image-tag subgraph. On the other hand, in the extreme case when $\lambda = 0$, no random walk will diffuse on the image-tag subgraph, and it only diffuses on the image similarity graph. In the intermediate case, when λ is relatively large, the diffusion on the image-tag subgraph is faster than the diffusion on the image similarity subgraph. As a result, the random walk will depend more on the image-tag information. If λ is relatively small, the results will depend more on the image visual information. In Section IV, we will give a detailed analysis on the impact of parameter λ.

With the transition probabilistic matrix $\mathbf{P}$ defined in (4), we can now perform the random walk on the hybrid graph: we calculate the probability of transition from node $\mathcal{V}_i$ to node $\mathcal{V}_j$ as

$$P_{t|0}(j|i) = [\mathbf{P}^t]_{ij}.^1 \quad (5)$$

1

The random walk sums the probabilities of all paths of length t between the two nodes. It gives a measure of the *volume of paths* between these two nodes; if there are many paths, the transition probability will be higher [9]. The larger the transition probability $P_{t|0}(j|i)$ is, the more the node $\mathcal{V}_j$ is similar to the node $\mathcal{V}_i$.

Since the image dataset is very large, computing the matrix multiplication $\mathbf{P}^t$ is infeasible. Hence, we compute the random walk in an efficient way as follows. If we want to start a random walk at node $\mathcal{V}_i$, we employ a row vector v_i with a unit entry at node $\mathcal{V}_i$, and then calculate the transition probability to node $\mathcal{V}_j$ as

$$P_{t|0}(j|i) = [(((v_i\mathbf{P})\mathbf{P})\ldots)\mathbf{P}]_j \quad (6)$$

where t controls the number of walk steps. This is very efficient since the matrix operations are quite sparse.

With the hybrid graph and the random walk model, similar to [9], we can then apply our framework to several application areas, including the following.

- **Image-to-image retrieval**. Given an image, find relevant images based on visual information and tags. The relevant documents should be ranked highly regardless of whether they are adjacent to the original image in the hybrid graph.
- **Image-to-tag suggestion**. This is also called image annotation. Given an image, find related tags that have semantic relations to the contents of this image.
- **Tag-to-image retrieval**. Given a tag, find a ranked list of images related to this tag. This is more like the text-based image retrieval.
- **Tag-to-tag suggestion**. Given a tag, suggest some other relevant tags to this tag. This is also known as tag recommendation problem.

In Section IV, we will show the performance of our model on the first three applications. We will not show the experiments of the tag recommendation application since it is beyond the scope of this paper.

D. Pseudo Relevance Feedback

Relevance feedback is an effective scheme to bridge the gap between high-level semantics and low-level features in content-based image retrieval. However, it involves the user interaction in the retrieval processes, which is infeasible in some retrieval applications. *Pseudo relevance feedback* provides an effective method for automatic local analysis. It automates the manual part of relevance feedback, so that users would obtain improved retrieval performance without an extended interaction.

Taking advantage of the proposed random work algorithm, our proposed framework can be naturally extended to pseudo relevance feedback. Consider the image-to-image retrieval example: Given an image, we first conduct a round of random walk, assuming that the top ranked images are relevant; then conduct another round of random walk, using the original image and the top ranked images. Actually, the top ranked images are used to make an expansion of the original image. We then re-rank all the images based on the expanded image set. The detailed algorithm for pseudo relevance feedback is summarized in Algorithm 1.

Algorithm 1: Pseudo Relevance Feedback Algorithm

1) Given the query node $\mathcal{V}_i$, form a $1 \times n$ vector v (n is the total number of nodes), with the ith entry equal to 1 while other entries equal to 0.
2) Perform a t-step random walk and get a new vector $v^* = v\mathbf{P}^t$.
3) Get the top-L nodes with the highest values in vector v^* (notice that, in the image-to-image retrieval task, the top-L nodes are in the image node set $\mathcal{D}$ while in the image-to-tag task or image annotation task, the top-L nodes are in the tag node set $\mathcal{T}$).
4) Form a new $1 \times n$ vector v' with the ith entry equal to 1, the entries represent the top-L nodes being equal to 1, and other entries being equal to 0.
5) Conduct a new t-step random walk and get the results $v^* = v'\mathbf{P}^\mathbf{t}$. Rank the vector v^* as the retrieval results.

In the above algorithm, we only conduct 1-round feedback. Actually, this algorithm can also run multiple feedback rounds.

IV. Experimental Analysis

In this section, we evaluate our proposed framework using the Flickr[2] dataset on content-based image retrieval, text-based image retrieval and image annotation problems.

A. Data Description

Flickr is an image hosting Web site and online community platform. It creates a popular platform for users to share per-

[1]Before the start of any random walks, we will normalize the probability $P_{t+1|t}(j|i)$ by $\sum_j P_{t+1|t}(j|i)$ to make sure that P is the transition probabilistic matrix.

[2]http://www.flickr.com.

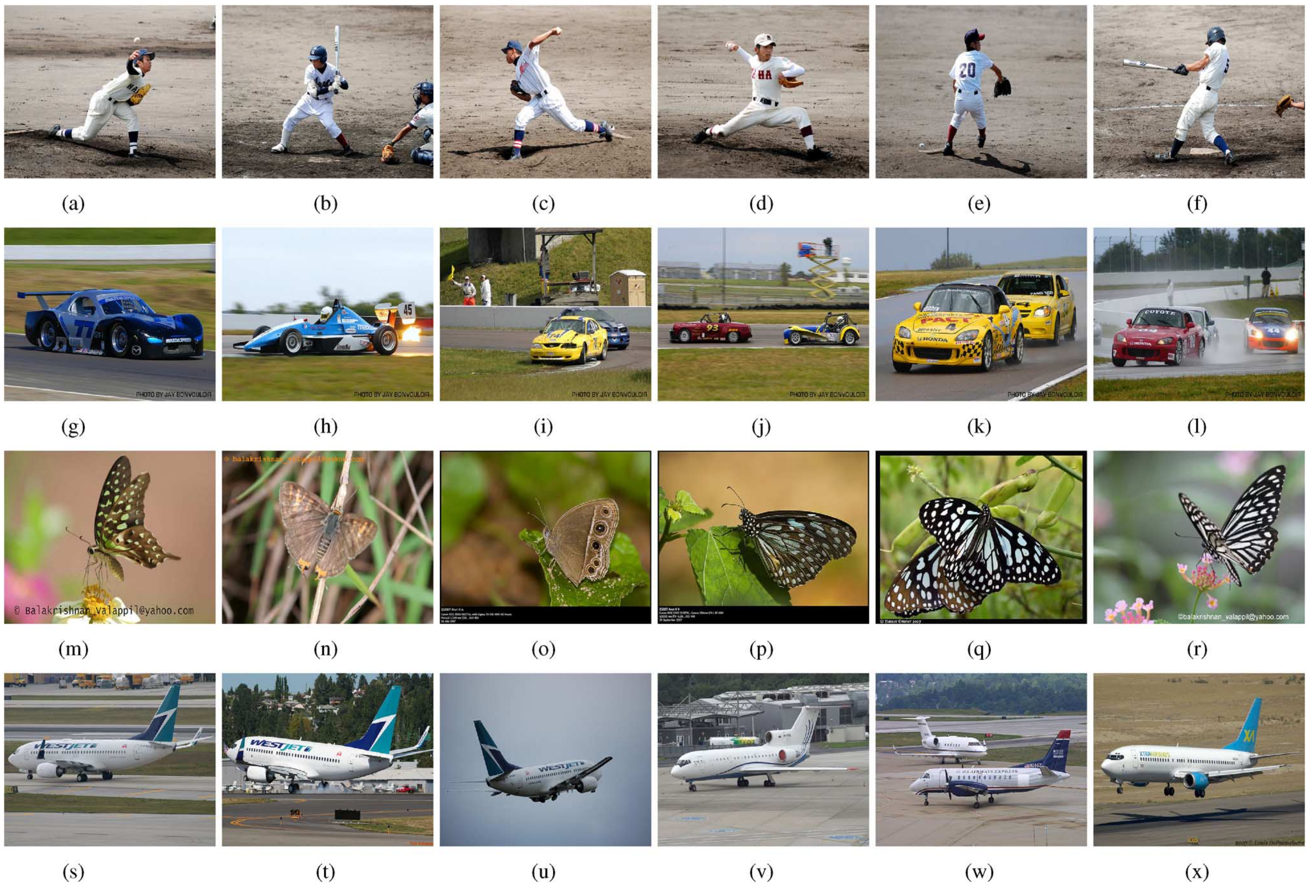


Fig. 3. Examples for CBIR. (a) Query Image 1. (b) Rank 1. (c) Rank 2. (d) Rank 3. (e) Rank 4. (f) Rank 5. (g) Query Image 2. (h) Rank 1. (i) Rank 2. (j) Rank 3. (k) Rank 4. (l) Rank 5. (m) Query Image 3. (n) Rank 1. (o) Rank 2. (p) Rank 3. (q) Rank 4. (r) Rank 5. (s) Query Image 4. (t) Rank 1. (u) Rank 2. (v) Rank 3. (w) Rank 4. (x) Rank 5.

sonal photographs, tag photographs, and communicate with other users. As of November 2007, it claims to host more than 2 billion images [3]. Hence, Flickr is an ideal source for the investigation of image-related research.

In this paper, we randomly sample 597 332 images which span from January 1, 2007 to December 31, 2007. For each image, we record the image files and the associated tags. Totally, we find 566 293 unique tags and 4 929 093 edges (tag assignments) between images and tags, which indicates that on average, each image has been associated with 8.25 tags.

B. Parameter Discussions

In addition to the fusion parameter λ, we need to set another two parameters: the parameter k for building kNN image similarity graph and the parameter t for the random walk steps.

For the parameter k, as suggested in [46], normally a small value of k performs well in practice. In this paper, we set $k = 40$ empirically in all of the experiments, which indicates that in the image similarity subgraph, every image has 40 most similar neighbors. Hence, the outdegree for every image node in the image similarity subgraph is 40.

The parameter t determines the resolution of the Markov random walk. If we choose a large enough t, the random walk will turn into the stationary distribution, where the final results depend more on the graph structure with little information about the query node preserved. On the other hand, a short walk preserves information about the starting node at a fine scale. Since we wish to preserve the information about the query node, a relatively small t is chosen in order to be far away from the stationary distribution. In this paper, we set $t = 10$ in all of our experiments.

For the parameter $\lambda \in [0, 1]$, it smoothly fuses the image-tag information with the image visual information, and directly controls how much the image-tag information should be trusted other than the image visual information. We will discuss the impact of this parameter in the three different applications considered in this research.

C. Content-Based Image Retrieval

In the content-based image retrieval (CBIR) tasks, we start the random walk at an image node. After a t-step random walk, we retrieve the top-ranked images as the retrieval results.

In Fig. 3, we perform four image retrievals with parameter $\lambda = 0.7$. (This is the best setting based on our empirical analysis, and we will discuss the impact of λ later in this section.) The first retrieval is based on the image plotted in Fig. 3(a), which is a picture depicting a baseball player. Fig. 3(b)–(f) shows the top-5 images returned by our method. We can observe that these results are all semantically related to the original picture. Fig. 3(g), (m), and (s) are another three examples, with Fig. 3(h)–(I), (n)–(r), and (t)–(x) as the retrieval results, respectively. The results show the excellent performance of our

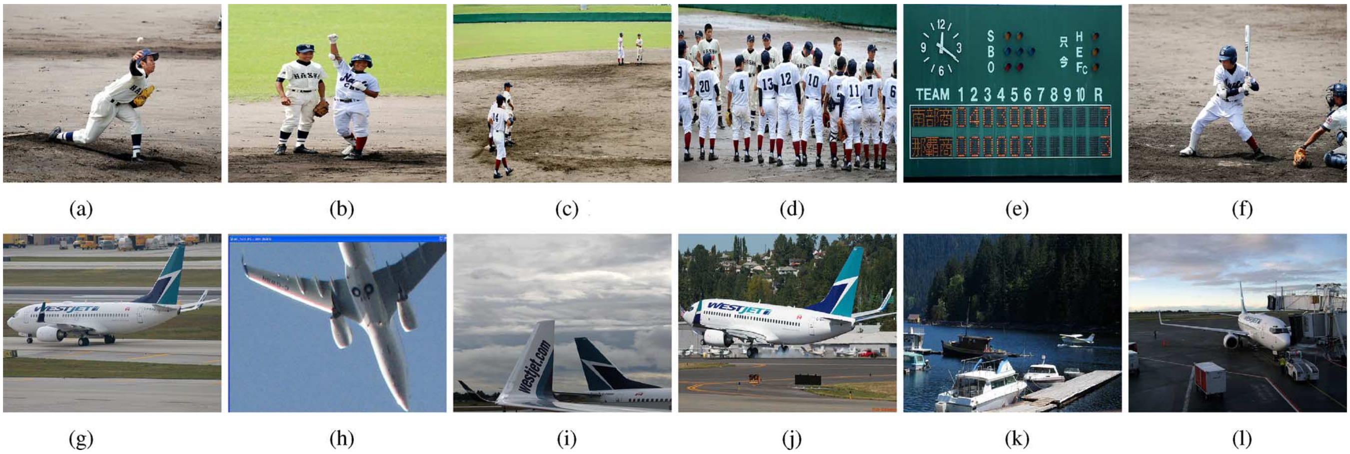


Fig. 4. Examples for RWIT method. (a) Query Image 1. (b) Rank 1. (c) Rank 2. (d) Rank 3. (e) Rank 4. (f) Rank 5. (g) Query Image 4. (h) Rank 1. (i) Rank 2. (j) Rank 3. (k) Rank 4. (l) Rank 5.

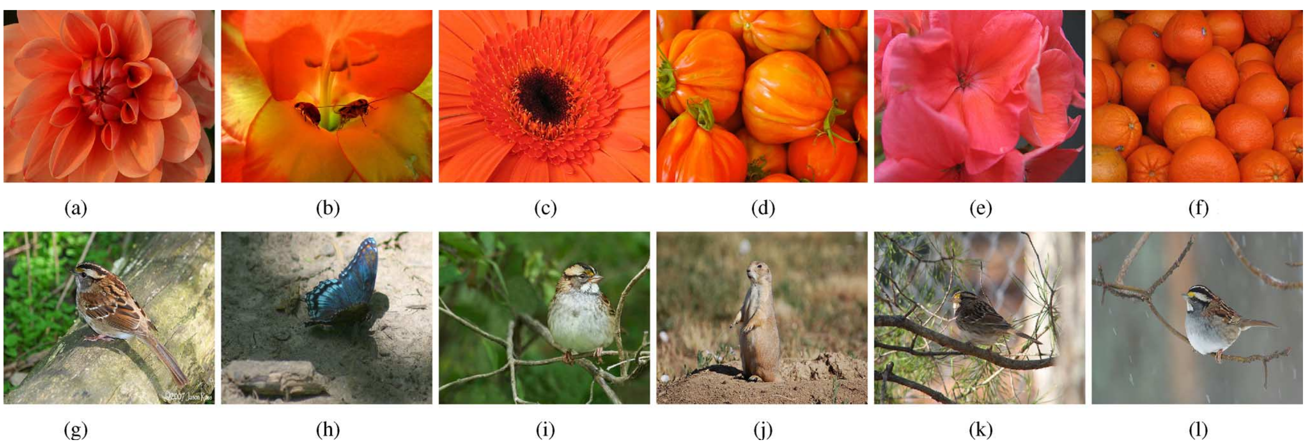


Fig. 5. Two failed examples for our method in CBIR. (a) Query Image 1. (b) Rank 1. (c) Rank 2. (d) Rank 3. (e) Rank 4. (f) Rank 5. (g) Query Image 2. (h) Rank 1. (i) Rank 2. (j) Rank 3. (k) Rank 4. (l) Rank 5.

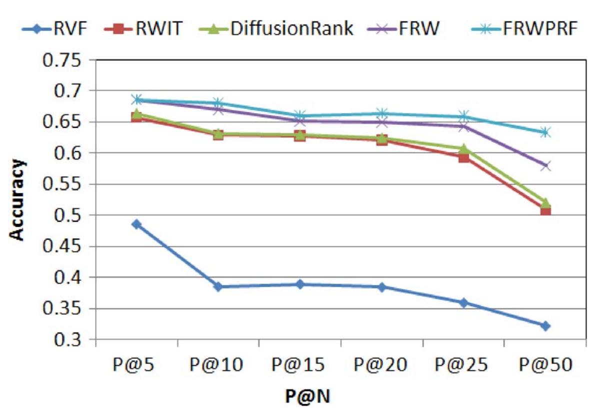


Fig. 6. P@N comparisons in CBIR.

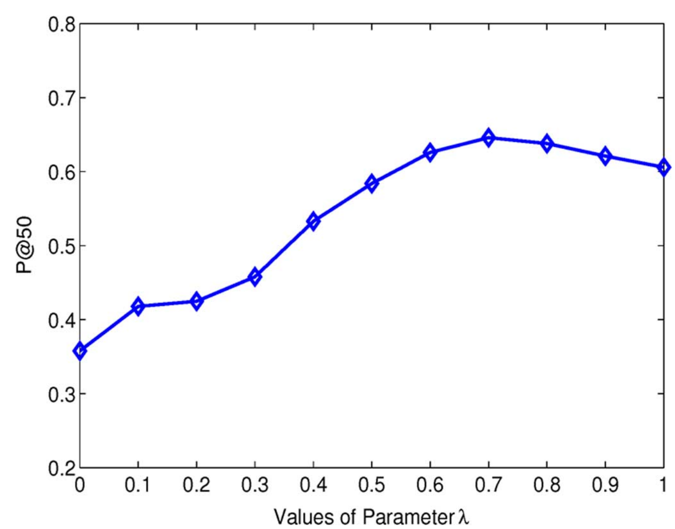


Fig. 7. Impact of parameter λ in CBIR.

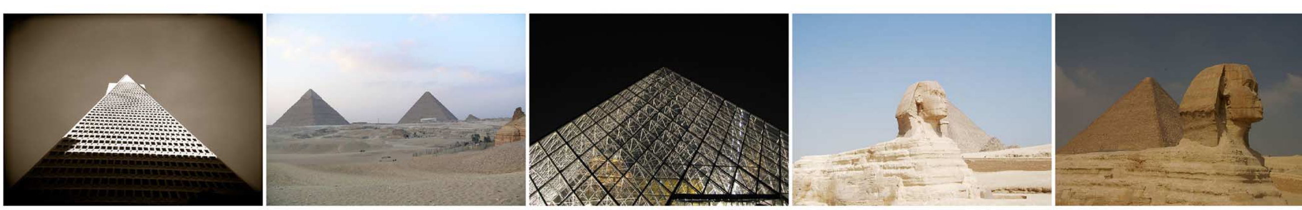
approach. We also list some of the results in Fig. 4 which are generated from the RWIT method proposed in [9]. We can see that our method can generate more reasonable results than the RWIT method. We also notice that in some cases, our algorithm cannot generate satisfactory results. Two such examples are illustrated in Fig. 5. For the first example, we can see that the query image (flower) and the recommended images have very similar color and edge distributions. This is because of that the extracted visual features mainly take account of the global color and edge distributions. Although the methods using local features [22], [42], [24] can alleviate this problem, it requires

(a)

(b)

(c)

(d)



(e)

Fig. 8. Examples of top images using text query. (a) Rainbow. (b) Grand Canyon. (c) Fireworks. (d) Basketball. (e) Pyramid.

high computational power and large storage space to calculate the local feature descriptors, especially for the Flickr photos with relatively large size. The second example also shows the similar problem.

In order to show the performance improvement of our approaches, we then compare our fusion by random walk (FRW) method and fusion by random walk with pseudo relevance feedback (FRWPRF) method with another three methods.

1) **RVF**: this method is a baseline method, which is purely based on image visual features. For every query image, we retrieve the top-N images using the similarity calculation function defined in (1). We call this method retrieval by visual feature (RVF) method.
2) **RWIT**: this method is based on the forward random walk model with self-transitions on the image-tag bipartite graph which is proposed in [9]. For every query image, we start the random walk at the query image node on the image-tag bipartite graph, and retrieve the top images as the results. We call this method random walk using image-tag (RWIT) relationships.
3) **DiffusionRank**: this method is a random walk method based on heat diffusion phenomenon which is proposed in [41]. For every query image, we start the heat diffusion process on the image-tag bipartite graph, and retrieve the top images with the largest heat values as the results. We call this method DiffusionRank.

In order to evaluate these methods, we use the metrics Precision@N. We select a set of 200 testing query images, and ask a panel of three experts to measure the relevance between the testing query images and the retrieved images. The Precision@N is defined as

$$\mathbf{P}@N = \frac{1}{|QI|}\sum_{i \in QI}\frac{R_i}{N} \tag{7}$$

(a)

(b)

Fig. 9. Examples of top images using text query by RWIT method. (a) Basketball. (b) Pyramid.

where the set QI contains all the testing query images, $|QI|$ is the number of the testing query images, R_i refers to the number of relevant images retrieved as to the ith testing query image, and N is the number of top images retrieved for every testing query image.

The comparison results are shown in Fig. 6. We can observe that our method FRW (with $\lambda = 0.7$) performs much better than the methods RVF, RWIT, and DiffusionRank. If we incorporate the pseudo relevance feedback algorithm proposed in Algorithm 1, the performance is further improved (in FRWPRF, we use the top-5 results as the feedback images). This shows the promising future of our proposed framework.

Parameter λ balances the information from image visual features and image-tag information. It takes advantages of these two types of information. If $\lambda = 1$, we only utilize the information from the image-tag bipartite graph; for $\lambda = 0$, we only mine the information from the image similarity graph. In other cases, we fuse these two sources together for the image retrieval tasks. To investigate the impact of parameter λ, we choose different values of λ to evaluate our FRW method. Fig. 7 plots the trend of the P@50 changing with parameter λ.

We can conclude that the value of λ affects the retrieval results significantly. This indicates that fusing these two sources will not always generate the best performance. We need to manually choose an appropriate value to avoid overtuning the parameters. Another interesting observation is when following the increase of the value of λ, the value of P@50 first increases, but when $\lambda > 0.7$, the value of P@50 starts to drop. This phenomenon demonstrates in most cases, low-level visual features contain less information than textual tags (that is why the optimal value of lambda is closer to 1 than to 0), but a combination of both sources of information usually achieves better results than using only one of them (that is why the optimal value of lambda is less than 1).

D. Text-Based Image Retrieval

In the text-based image retrieval (TBIR), we start the random walk at a tag node. After a t-step random walk, we select the top-ranked images as the retrieval results.

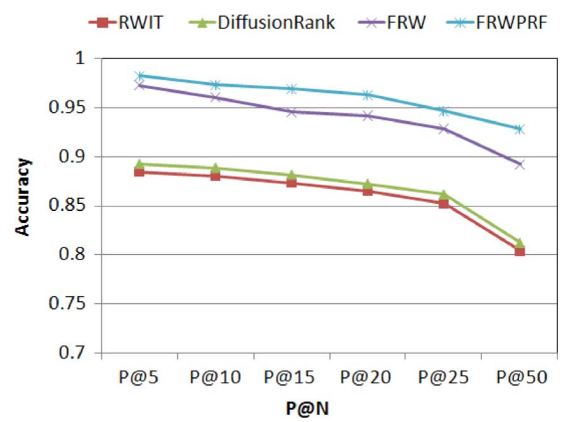


Fig. 10. P@N comparisons in TBIR.

Fig. 8 shows five TBIR examples (with $\lambda = 0.7$). The queries are "Rainbow", "Grand Canyon", "Fireworks", "Basketball", and "Pyramid", respectively. From the retrieved top-5 results, we can observe that our method performs very well. We also list some of the results generated by RWIT in Fig. 9 for comparison. We create a set of 200 queries, and compare our FRW method and FRWPRF method in terms of TBIR with RWIT and DiffusionRank methods, which only utilize the image-tag relationships for retrieval. Fig. 10 describes the comparison results. We find that both FRW and the relevance feedback method FRWPRF perform much better than RWIT and DiffusionRank method. The parameter λ also play an important role in TBIR. Basically, it shares the same trend with CBIR, and the optimal value of λ is also around 0.7.

E. Image Annotation

Automated image annotation has been an active and challenging research topic in computer vision and pattern recognition for years. Automated image annotation is essential to make huge unlabeled digital photos indexable by existing text-based indexing and search solutions. In general, an image annotation task consists to assign a set of semantic tags or labels to a novel image based on some models learned from certain training data. Conventional image annotation approaches often attempt

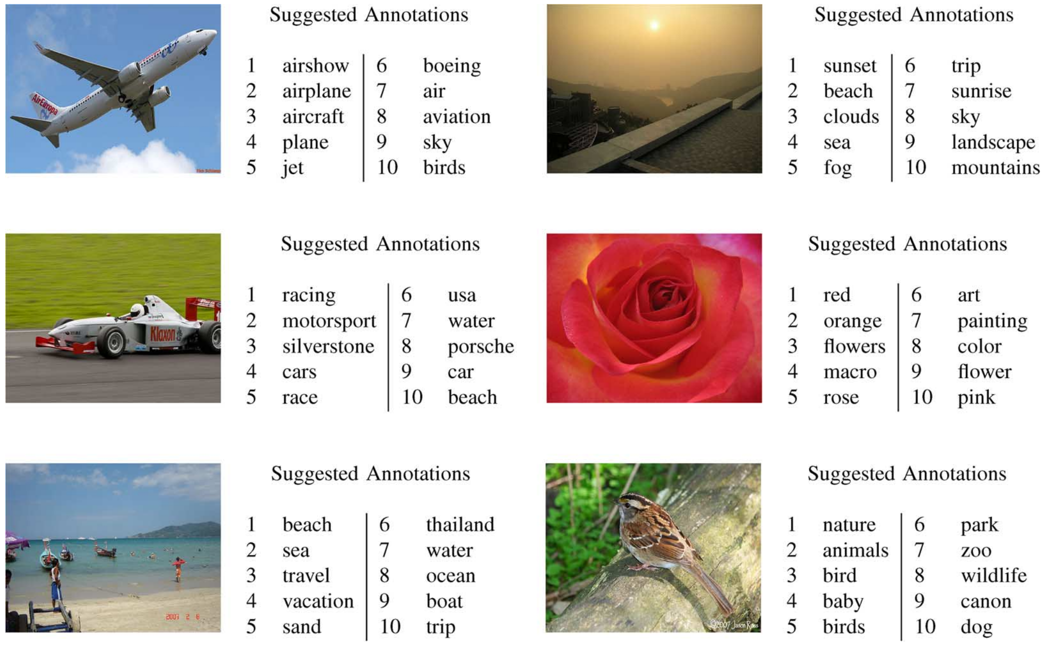


Fig. 11. Examples of image annotations.

TABLE I
P@N IN AUTOMATED IMAGE ANNOTATION

	P@1	**P@2**	**P@3**	**P@4**	**P@5**	**P@6**	**P@7**	**P@8**
FRW	0.495	0.431	0.428	0.388	0.366	0.314	0.301	0.296

to detect semantic concepts with a collection of human-labeled training images. Due to the long-standing challenge of object recognition, such approaches, though working reasonably well for small-sized testbeds, often perform poorly on large dataset in the real world. Besides, it is often expensive and time-consuming to collect the training data.

In addition to the success in image retrieval, our FRM framework also provides a natural, effective, and efficient solution for automated image annotation tasks. For every new image, we first extract a 297-dimensional feature vector through the method described in Section III-A. Then, we find the top-k similar images using (1) ($k = 40$ in our experiments), and link the new image to the top-k images in the hybrid graph. Finally, we start the random walk at this newly built image node, and return the top-N tags as the annotations to this image.

Fig. 11 gives six examples to demonstrate the qualitative performance of the annotation results by our framework (with $\lambda = 0.2$). We select a set of 50 images as the testing images for automated image annotation. The image annotation accuracy is shown in Table I, which demonstrates a very competitive result since automated image annotation is a very challenging problem. We also observe that the trend of accuracy changing with parameter λ is not similar to the one in CBIR and TBIR. As shown in Fig. 12, we can observe that the optimal value of parameter λ is around 0.2. This is because at the beginning of the random walk, the starting image does not have any links connected to the tag nodes; hence, we need to trust more on the image similarity subgraph. Otherwise, we cannot generate accurate image annotations, and the overall precision will suffer.

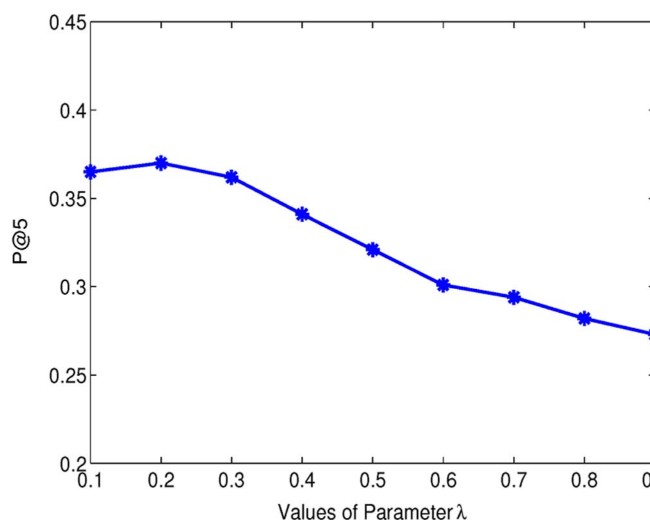


Fig. 12. Impact of parameter λ in image annotation.

V. Conclusions and Future Work

In this paper, we present a novel framework for several image retrieval tasks based on Markov random walk. The proposed framework bridges the semantic gap existing between visual contents and textual tags in a simple but efficient way. We do not need to train any learning function or any training data; hence, our method can be easily adapted to very large datasets. Finally, the experimental results on a large Flickr dataset show the effectiveness of our approach.

In the future, we plan to incorporate more information into our proposed framework. Specifically, we only utilize the image contents and the image tags information in this paper. Actually, there are lots of metadata on Flickr websites, such as the social network information among users and the image notes information, which can also be employed to improve the retrieval performance. Moreover, we need to design a more flexible model to include all these information pieces. Another problem worthy of investigation is to develop other Markov random walk models. Instead of using the "forward" random walk mode in this paper, we can also try the models like the "backward" model, and compare their performance. Another problem worthy of investigation is that how the amount and quality of tags could affect the performance of our method.

ACKNOWLEDGMENT

The authors also would like to thank the reviewers and associate editor for their helpful comments.

REFERENCES

[1] M. Ames and M. Naaman, "Why we tag: Motivations for annotation in mobile and online media," in *Proc. CHI'07*, San Jose, CA, 2007, pp. 971–980.

[2] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *J. ACM*, vol. 45, no. 6, pp. 891–923, 1998.

[3] E. Auchard, "Flickr to map the world's latest photo hotspots," *Reuters*, 2007.

[4] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, and M. Aly, "Video suggestion and discovery for youtube: Taking random walks through the view graph," in *Proc. WWW'08*, Beijing, China, 2008, pp. 895–904.

[5] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," *J. Mach. Learn. Res.*, vol. 3, pp. 1107–1135, 2003.

[6] D. M. Blei and M. I. Jordan, "Modeling annotated data," in *Proc. SIGIR'03*, Toronto, ON, Canada, 2003, pp. 127–134.

[7] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[8] O. Chum, M. Perdoch, and J. Matas, "Geometric min-hashing: Finding a (thick) needle in a haystack," in *Proc. CVPR'09*, 2009, pp. 17–24.

[9] N. Craswell and M. Szummer, "Random walks on the click graph," in *Proc. SIGIR'07*, Amsterdam, The Netherlands, 2007, pp. 239–246.

[10] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, pp. 1–60, 2008.

[11] C. Djeraba, "Association and content-based retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 1, pp. 118–135, Jan.–Feb. 2003.

[12] P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," in *Proc. 7th Eur. Conf. Computer Vision—Part IV (ECCV'02)*, London, U.K., 2002, pp. 97–112, Springer-Verlag.

[13] N. Eiron, K. S. McCurley, and J. A. Tomlin, "Ranking the web frontier," in *Proc. WWW'04*, New York, 2004, pp. 309–318.

[14] ESP. [Online]. Available: http://www.espgame.org.

[15] J. Fan, Y. Gao, H. Luo, and R. Jain, "Mining multilevel image semantics via hierarchical classification," *IEEE Trans. Multimedia*, vol. 10, no. 2, pp. 167–187, Feb. 2008.

[16] J. Fan, Y. Gao, H. Luo, and G. Xu, "Automatic image annotation by using concept-sensitive salient objects for image content representation," in *Proc. SIGIR'04*, Sheffield, U.K., 2004, pp. 361–368.

[17] Flickr. [Online]. Available: http://www.fickr.com.

[18] A. Ghoshal, P. Ircing, and S. Khudanpur, "Hidden Markov models for automatic annotation and content-based retrieval of images and video," in *Proc. SIGIR'05*, Salvador, Brazil, 2005, pp. 544–551.

[19] A. Grigorova, F. G. B. D. Natale, C. K. Dagli, and T. S. Huang, "Content-based image retrieval by feature adaptation and relevance feedback," *IEEE Trans. Multimedia*, vol. 9, no. 6, pp. 1183–1192, Oct. 2007.

[20] W. H. Hsu, L. S. Kennedy, and S.-F. Chang, "Video search reranking through random walk over document-level context graph," in *Proc. MM'07*, Augsburg, Germany, 2007, pp. 971–980.

[21] J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models," in *Proc. SIGIR'03*, Toronto, ON, Canada, 2003, pp. 119–126.

[22] Y. Jing and S. Baluja, "PageRank for product image search," in *Proc. WWW'08*, Beijing, China, 2008, pp. 307–316.

[23] R. Krishnapuram, S. Medasani, S.-H. Jung, Y. Choi, and R. Balasubramaniam, "Content-based image retrieval based on a fuzzy approach," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 10, pp. 1185–1199, Oct. 2004.

[24] Y.-H. Kuo, K.-T. Chen, C.-H. Chiang, and W. H. Hsu, "Query expansion for hash-based image object retrieval," in *Proc. MM'09*, Beijing, China, 2009, pp. 65–74.

[25] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Comput.*, vol. 42, no. 3, pp. 300–311, Mar. 1993.

[26] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, no. 1, pp. 1–19, 2006.

[27] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[28] H. Ma, I. King, and M. R. Lyu, "Effective missing data prediction for collaborative filtering," in *Proc. SIGIR'07*, Amsterdam, The Netherlands, 2007, pp. 39–46.

[29] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with social trust ensemble," in *Proc. SIGIR'09*, Boston, MA, 2009, pp. 203–210.

[30] H. Ma, H. Yang, M. R. Lyu, and I. King, "SoRec: Social recommendation using probabilistic matrix factorization," in *Proc. CIKM'08*, Napa Valley, CA, 2008, pp. 931–940.

[31] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[32] L. Page, S. Brin, R. Motwani, and T. Winograd, The Pagerank Citation Ranking: Bringing Order to the Web, Tech. Rep. Paper, 1999, SIDLWP-1999-0120 (version of 11/11/1999).

[33] K. Pearson, "The problem of the random walk," *Nature*, vol. 72, pp. 294–294, 1905.

[34] G.-J. Qi, X.-S. Hua, and H.-J. Zhang, "Learning semantic distance from community-tagged media collection," in *Proc. MM'09*, Beijing, China, 2009, pp. 243–252.

[35] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool in interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 644–655, Sep. 1998.

[36] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[37] J. Tang, H. Li, G.-J. Qi, and T.-S. Chua, "Image annotation by graph-based inference with integrated multiple/single instance representations," *IEEE Trans. Multimedia*, vol. 12, no. 2, pp. 131–141, Feb. 2010.

[38] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proc. MM'01*, Ottawa, ON, Canada, 2001, pp. 107–118.

[39] C. Wang, L. Zhang, and H.-J. Zhang, "Learning to reduce the semantic gap in web image retrieval and annotation," in *Proc. SIGIR'08*, Singapore, 2008, pp. 355–362.

[40] L. Wu, X.-S. Hua, N. Yu, W.-Y. Ma, and S. Li, "Flickr distance," in *Proc. MM'08*, Vancouver, BC, Canada, 2008, pp. 31–40.

[41] H. Yang, I. King, and M. R. Lyu, "DiffusionRank: A possible penicillin for web spamming," in *Proc. SIGIR'07*, Amsterdam, The Netherlands, 2007, pp. 431–438.

[42] S. Zhang, Q. Tian, G. Hua, Q. Huang, and S. Li, "Descriptive visual words and visual phrases for image applications," in *Proc. MM'09*, Beijing, China, 2009, pp. 75–84.

[43] J. Zhu, S. C. Hoi, and M. R. Lyu, "Face annotation by transductive kernel fisher discriminant," *IEEE Trans. Multimedia*, vol. 10, no. 1, pp. 86–96, Jan. 2008.

[44] J. Zhu, S. C. Hoi, M. R. Lyu, and S. Yan, "Near-duplicate keyframe retrieval by nonrigid image matching," in *Proc. MM'08*, 2008, pp. 41–50.

[45] J. Zhu, S. C. Hoi, M. R. Lyu, and S. Yan, "Near-duplicate keyframe retrieval by semi-supervised learning and nonrigid image matching," *ACM Trans. Multimedia Comput., Commun. Appl.*, to be published.

[46] X. Zhu, "Semi-supervised learning with graphs," Ph.D. dissertation, Pittsburgh, PA, 2005, Chair-John Lafferty and Chair-Ronald Rosenfeld.

Hao Ma received the B.Eng. and M.Eng. degrees from the School of Information Science and Engineering at Central South University, Changsha, China, in 2002 and 2005, respectively, and the Ph.D. degree from the Computer Science and Engineering Department, The Chinese University of Hong Kong, Kowloon, in 2010.

He worked as a System Engineer in Intel Shanghai, Shanghai, China, before he joined CUHK as a Ph.D. student in November 2006. His research interests include information retrieval, data mining, machine learning, social network analysis, recommender systems, human computation, and social media analysis.

Jianke Zhu (M'09) received the B.S. degree in mechatronics and computer engineering from Beijing University of Chemical Technology, Beijing, China, the M.S. degree in electrical and electronics engineering from University of Macau, Taipa, Macau, and the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Kowloon.

He is currently a postdoc in the Computer Science and Engineering Department in The Chinese University of Hong Kong. His research interests include computer vision, machine learning, pattern recognition, and multimedia information retrieval.

Michael Rung-Tsong Lyu (F'04) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1981; the M.S. degree in computer engineering from the University of California, Santa Barbara, in 1985; and the Ph.D. degree in computer science from the University of California, Los Angeles, in 1988.

He was with the Jet Propulsion Laboratory as a Technical Staff Member from 1988 to 1990. From 1990 to 1992, he was with the Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, as an Assistant Professor. From 1992 to 1995, he was a Member of the Technical Staff in the applied research area of Bell Communications Research (Bellcore), Morristown, NJ. From 1995 to 1997, he was a Research Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ. In 1998, he joined The Chinese University of Hong Kong, Jowloon, where he is now a Professor in the Department of Computer Science and Engineering. He is also Founder and Director of the Video over InternEt and Wireless (VIEW) Technologies Laboratory. His research interests include software reliability engineering, distributed systems, fault-tolerant computing, mobile and sensor networks, Web technologies, multimedia information processing and retrieval, and machine learning. He has published 330 refereed journal and conference papers in these areas. He was the editor of two book volumes: *Software Fault Tolerance* (New York: Wiley, 1995) and *The Handbook of Software Reliability Engineering* (Piscataway, NJ: IEEE and New York: McGraw-Hill, 1996).

Dr. Lyu initiated the First International Symposium on Software Reliability Engineering (ISSRE) in 1990. He was the Program Chair for ISSRE 1996 and General Chair for ISSRE 2001. He was also PRDC 1999 Program Co-Chair, WWW10 Program Co-Chair, SRDS 2005 Program Co-Chair, PRDC 2005 General Co-Chair, ICEBE 2007 Program Co-Chair, and SCC 2010 Program Co-Chair. He will be the General Chair for DSN 2011 in Hong Kong. He was on the Editorial Board of the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON RELIABILITY, the *Journal of Information Science and Engineering*, and the *Wiley Software Testing, Verification & Reliability Journal*. Dr. Lyu is an AAAS Fellow and a Croucher Senior Research Fellow.

Irwin King (SM'08) received the B.Sc. degree in engineering and applied science from the California Institute of Technology, Pasadena, and the M.Sc. and Ph.D. degrees in computer science from the University of Southern California, Los Angeles.

He is with the Chinese University of Hong Kong, Kowloon. His research interests include machine learning, web intelligence, social computing, data mining, and multimedia information processing. In these research areas, he has over 200 technical publications in journals and conferences. In addition, he has contributed over 20 book chapters and edited volumes. Moreover, he has over 30 research and applied grants. One notable patented system he has developed is the VeriGuide System, which detects similar sentences and performs readability analysis of text-based documents in both English and Chinese to promote academic integrity and honesty.

Dr. King is an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS (TNN) and the IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE (CIM). He is a member of the Editorial Board of the *Open Information Systems Journal*, *Journal of Nonlinear Analysis and Applied Mathematics*, and *Neural Information ProcessingCLetters and Reviews Journal* (NIP-LR). He has also served as Special Issue Guest Editor for *Neurocomputing*, *International Journal of Intelligent Computing and Cybernetics* (IJICC), *Journal of Intelligent Information Systems* (JIIS), and *International Journal of Computational Intelligent Research* (IJCIR). He is a member of ACM, International Neural Network Society (INNS), and Asian Pacific Neural Network Assembly (APNNA). Currently, he is serving the Neural Network Technical Committee (NNTC) and the Data Mining Technical Committee under the IEEE Computational Intelligence Society (formerly the IEEE Neural Network Society). He is also a member of the Board of Governors of INNS and a Vice-President and Governing Board Member of APNNA.