# Social Visual Image Ranking for Web Image Search

Shaowei Liu[1], Peng Cui[1], Huanbo Luan[1,3], Wenwu Zhu[1], Shiqiang Yang[1], and Qi Tian[2]

[1] Beijing Key Laboratory of Networked Multimedia, Tsinghua University, China
[2] Computer Science Department, University of Texas at San Antonio, USA
[3] School of Computing, National University of Sinapore
liu-sw11@mails.tsinghua.edu.cn, cuip@mail.tsinghua.edu.cn,
luanhuanbo@gmail.com, {wwzhu,yangshq}@tsinghua.edu.cn, qitian@cs.utsa.edu

**Abstract.** Many research have been focusing on how to match the textual query with visual images and their surrounding texts or tags for Web image search. The returned results are often unsatisfactory due to their deviation from user intentions. In this paper, we propose a novel image ranking approach to web image search, in which we use social data from social media platform jointly with visual data to improve the relevance between returned images and user intentions (i.e., social relevance). Specifically, we propose a community-specific Social-Visual Ranking(SVR) algorithm to rerank the Web images by taking social relevance into account. Through extensive experiments, we demonstrated the importance of both visual factors and social factors, and the effectiveness and superiority of the social-visual ranking algorithm for Web image search.
**Keywords:** Social image search, Image reranking, Social relevance.

## 1 INTRODUCTION

Fundamentally, user intention plays an important role in image search. Most of traditional image search engines represent user intentions with textual query. Thus, a lot of existing research work focuses on improving the relevance between the textual query and visual images. However, there exists semantic gap between user intention and textual query. Let's take the query "jaguar" as an example, as shown in Fig.1. Different users have different intentions when inputting the query "jaguar". Some are expecting leopard images, while others are expecting automobile images. This scenario is quite common, particularly for queries with heterogeneous concepts or general (non-specific) concepts. This raises a fundamental but yet unsolved problem in Web image search: how to understand user intentions when users conducting image search?

Today user interests is mostly used to understand user intentions. For the instance in last paragraph, if we have the knowledge that the user is interested in animals, we can infer that he is likely to want the images about leopards when he searches "jaguar". In the past years, interest analysis is very difficult due to the lack of personal data. With the development of social media platforms, such as Flickr and Facebook, the way people can get social data has been changed: users' profiles, interests and their favorite images are exposed online and open to public, which are crucial information sources to implicitly understand user interests.In this paper, we exploit social data to assist image search, aiming to improve the relevance between returned images and user interests, which is termed as *Social Relevance*.

By considering social relevance and visual relevance comprehensively, we can understand user intention better, thereby improving the performance of our image ranking approach. However, the combination faces the following challenges:

(1) **Social data sparseness**. In social media platform, most users only possess a small number of favored images, from which it is difficult to discover user intentions. With the hypothesis that users in the same community share similar interests, a community-specific method is more practical and effective than a user-specific method.

(2) **The tradeoff between social relevance and visual relevance**. Although social relevance may guarantee the interest of returned images for the user, the quality and representativeness of images, cannot be ignored. Both of which are necessary for good search results. Thus, both social relevance and visual relevance are needed to be addressed and subtly balanced.

(3) **Complex factors**. To generate the final image ranking, one needs to consider the user query, returned images from current search engines, and many complex social factors derived from social media platforms. How to integrate these heterogeneous factors in an effective and efficient way is quite challenging.

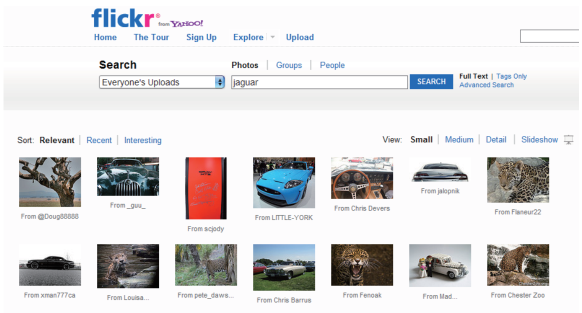


**Fig. 1.** The results returned by Flickr for the query "jaguar", recorded on April, 10th, 2012.

To address the above problems, we propose a novel community-specific Social-Visual Ranking (SVR) algorithm to rerank the Web images returned by current image search engines. More specifically, in SVR, given the preliminary image search results and the user's Flickr ID, we will use group information in social platform and visual contents of the images to rerank the Web images for a group that the user belongs to, which is termed as *the user's membership group*. In SVR, user interests and textual query are both utilized to predict user intention. The SVR algorithm is implemented by PageRank over a hybrid image link graph, which is the combination of an image social-link graph and an image visual-link graph. Through SVR, the Web images are reranked according to their interests to the users while maintaining high visual quality and representativeness for the query.

The contributions of our proposed approach are highlighted as follows:

1) We propose a novel image ranking method for by combining the information in social media platforms and traditional image search engines to address the user intention understanding problem in Web image search, which is of ample significance to improve image search performances.

2) We propose a community-specific Social-Visual Ranking algorithm to rerank Web images according to their social relevance and visual relevance. In this algorithm, complex social and visual factors are effectively and efficiently incorporated by hybrid image link graph, and more factors can be naturally enriched.

3) We have conducted intensive experiments, indicated the importance of both visual factors and social factors, and demonstrated the advantages of social-visual ranking algorithms for Web image search. Except image search, our algorithm can also be straightforwardly applied in other related areas, such as product recommendation and personalized advertisement.

The rest of the paper is organized as follows. We introduce some related works in Section 2. Image link graph generation and image ranking is presented in Section 3. Section 4 presents the details and analysis of our experiments. Finally, Section 5 concludes the paper.

## 2 RELATED WORK

Aiming at improving the visual relevance, a series of methods are proposed based on incorporating visual factors into image ranking [2, 8]. An essential problem in these methods is to measure the visual similarity[5]. As an effective approach, VisualRank[4] determines the visual similarity by the number of shared SIFT features[1]. After a similarity based image link graph was generated, an iterative computation similar to PageRank[11] is utilized to rerank the images. VisualRank obtains a better performance than text-based image search in the measurement of relevance for queries with homogeneous visual concepts. However, for queries with heterogeneous visual concepts, VisualRank does not work well[9].

With the development of social media platform, the concept of social image retrieval was proposed, which brings more information and challenges to us[14]. Most of works in social image search focus on tags [15, 7, 6]. However, the quality of recommendation is based on the technique of tag annotation[13], which is not mature enough. Overall, understanding user intention is significant but challengeable in social media platform. Many social media sites such as Flickr offer millions of groups for users to share images with others. There are tons of works based on improving the user experience [12]. Group information is an efficient wav to estimate user interests.

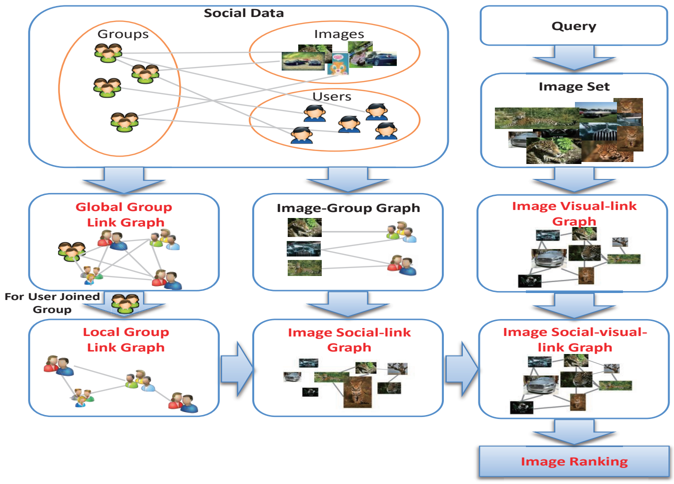


**Fig. 2.** The framework of our approach. Major intermediate results and final target are marked in red.

## 3 SOCIAL-VISUAL RERANKING

Fig.2 illustrates the framework of our Social-Visual ranking algorithm. In our approach, a random walk model based on PageRank[11] is utilized for image ranking. The weight $p(I_i, I_j)$ of the link from image $I_i$ to image $I_j$ represents the probability that a user will jump to $I_j$ after viewing $I_i$. This procedure can be considered in both social factor and visual factor. From the social point of view, if $I_i$'s group $G_p$ is similar to $I_j$'s group $G_q$, the probability of user's jump from $I_i$ to $I_j$ will be high. From the visual point of view, a user may be attracted by some visual contents of $I_i$ and then decide to view $I_j$ which also contains these contents. As a result, these two factors will both have significant effects in image ranking. Thus, we define our image link graph as the linear combination of the visual-link graph and the social-link graph. i.e.,

$$P_G = \alpha \cdot P_G^S + (1 - \alpha) \cdot P^V \tag{1}$$

where $P_G$ is the adjacency matrix of the hybrid image link graph. $P_G^S$ is the matrix for image social-link and $P^V$ is the matrix of image visual-link graph. $\alpha$ is a parameter to balance these factors. The estimation of $\alpha$ will be discussed in Section 4. In this equation, $P_G$ and $P_G^S$ are relevant to the user's membership group $G$. Therefore they have a subscript as '$G$'. The symbols with the subscript '$G$' in our algorithm have the same meaning.

### 3.1 Image Social-link Graph

**Global group ranking** First, a global group link graph is generated in preprocessing phase of our algorithm based on group similarity. Group similarity in user interests can be measured by the overlap of user sets and data sets, which is defined as:

$$S(G_u, G_v) = \lambda \cdot overlap(\mathcal{M}_u, \mathcal{M}_v) + (1 - \lambda) overlap(\mathcal{I}_u, \mathcal{I}_v) \tag{2}$$

where $\mathcal{M}_i$ is the user set of group $G_u$ and $\mathcal{I}_u$ is the image set of group $G_u$. $\lambda$ is a parameter to balance the user factor and the image factor. The overlap of $\mathcal{M}_i$ and $\mathcal{M}_j$ can be described as the Jaccard distance:

$$overlap(\mathcal{M}_u, \mathcal{M}_v) = \frac{\mathcal{M}_u \cap \mathcal{M}_v}{\mathcal{M}_u \cup \mathcal{M}_v} \tag{3}$$

so is the overlap of $\mathcal{I}_u$ and $\mathcal{I}_v$.

After the pair-wised group similarities are computed, the iterative computation based on PageRank can be utilized to evaluate the centrality of the groups:

$$gr = d \cdot S \cdot gr + (1 - d) e_0, e_0 = \left[\frac{1}{N_G}\right]_{N_G \times 1} \tag{4}$$

where $S$ is a column-normalized matrix constructed by $S(G_u, G_v)$. $N_G$ is the number of groups. d is the probability for user to visit the images along the graph links rather than randomly.

**Local group link graph** Local group link graph can be generated based on social strength of pairwise groups. Social strength of group $G_u$ and group $G_v$ for the given membership group $G$, represented as $T_G(G_u, G_v)$, describes the correlation between $G_u$ and $G_v$ with respect to $G$'s interests. In other words, $T_G(G_u, G_v)$ denotes the probability that an user in $G$ will jump to the images of $G_v$ after viewing the images of $G_u$.

The group similarity $S(G_u, G_v)$ can represent the degree that $G_u$ recommend $G_v$ to $G$. If users in $G$ are interested in images in $G_u$ and $G_u$ recommend $G_v$ to $G$, then users in $G$ may also be interested in images in $G_v$. Therefore, we can formulate the social strength $T_G(G_u, G_v)$:

$$T_G(G_u, G_v) = (S(G, G_u) + S(G, G_v)) \cdot S(G_u, G_v) \cdot f(gr(G_u)) \cdot f(gr(G_v)) \tag{5}$$

where $f(gr(G_u))$ is a function of the group rank value of $G_u$ in the rank vector calculated in Eq.4. It denotes the weight of group importance. In this paper, we just consider the basic form of power function, which is proved to be valid[16], i.e.:

$$f(x) = x^r \tag{6}$$

where $r$ is a parameter which will be estimated by experimental study.

**Image social-link graph** For images and groups, we first construct a basic image-group graph. The edge from an image to a group denotes the image belonging to the group:

$$A(I_i, G_u) = \begin{cases} 1 \ I_i \text{ belongs to } G_u \\ 0 \text{ otherwise} \end{cases} \tag{7}$$

Based on local group link graph and image-group graph, we can define the weight of the edge in image social-link graph as:

$$p_G^S(I_i, I_j) = \frac{Z_1}{(\sum_{u=1}^{N_G} A(I_i, G_u))(\sum_{u=1}^{N_G} A(I_j, G_u))} \cdot \sum_{u=1}^{N_G} \sum_{v=1}^{N_G} A(I_i, G_u) \cdot A(I_j, G_v) \cdot T(G_u, G_v) \tag{8}$$

where $Z_1$ is a column-normalization factor to normalize $\sum_j p_G^S(I_i, I_j)$ to 1. $p_G^S(I_i, I_j)$ denotes the probability that group $G$ will visit $I_j$ after viewing $I_i$ in social factor.

### 3.2 Image Visual-link Graph and Social-Visual Ranking

SIFT descriptors of the images are clustered into some visual words by a hierarchical visual vocabulary tree[10]. Then, an image can be regarded as a document including some words. The weight of the edge in visual image link graph can be defined as:

$$p^V(I_i, I_j) = \frac{C(I_i, I_j)}{\sum_i C(I_i, I_j)} \tag{9}$$

where $C(I_i, I_j)$ is the count of co-occurrence of visual words in image $I_i$ and $I_j$.

After two image link graphs are generated, hybrid image link graph can be constructed by Eq.1. Then, the iteration procedure based on PageRank can be formulated as:

$$r_G = d \cdot P_G \cdot r_G + (1 - d)e \tag{10}$$

where $d = 0.8$ as in Eq.4. $e$ is a parameter to describe the probability a user jumps to another image without links when he is tired of surfing by links. In our experiments, we have two choices of $e$:

$$e_1(i) = \frac{1}{N_I} \tag{11}$$

where $N_I$ is the number of images, and

$$e_G(i) = Z_2 \frac{\sum_{u=1}^{N_G} A(I_i, G_u) \cdot S(G, G_u)}{\sum_{u=1}^{N_G} A(I_i, G_u)} \tag{12}$$

where $Z_2$ is the factor to normalize the sum of $\sum e_G(i)$ to 1. These two cases of $e$ will be compared in our experiments.

# 4 EXPERIMENTS

## 4.1 Dataset and Settings

In this paper, we conduct experiments with data including images, groups, users, group-user relations and group-image relations from Flickr.com. 30 queries are collected and 1000 images are downloaded for each query by Flickr API. The selected queries includes:(1)Daily articles with no less than two different meanings, such as "apple", "jaguar" and "golf";(2) Natural scenery photos with multiple visual categories, such as "landscape", "scenery" and "hotel";(3)Living facilities with indoor and outdoor views, such as "restaurant" and "hotel";(4)Fashion products with different product types, such as "smart phone" and "dress".

In our experiment, we compare our algorithm SVR with other three image ranking methods: VisualRank(VR), SocialRank(SR) and Flickr search engine by relevance(FR) as baseline. Among them, VR is the special case for SVR when $\alpha = 0$, and SR is the special case for $\alpha = 1$.

## 4.2 Measurements

**Social relevance** Defined as the relevance to user intention, social relevance is an important measurement in our experiments. For a query, we randomly select $n$ testing pairs $(I_i, G_u)$ from the dataset, which means a group $G_u$ and an image $I_i$ belongs to this group. When a user in $G_u$ inputs a query, $I_i$ should be one of the images he wants to find. In another word, $I_i$ should get a high rank order in our algorithm. Therefore, we define a measurement called Average Rank($AR$) to reflect the degree to which we can capture user intentions:

$$AR = \frac{1}{|T|} \sum_{I_i \in T} rank(I_i) \tag{13}$$

where $T$ is the set of testing pairs. $rank(I_i)$ is the image $I_i$'s ranking order. In our experiments, we select 20 testing pairs for each query. The smaller the AR value, the better the algorithm performance.

**Visual relevance** All images in our dataset are labeled according to their visual relevance in 4 levels, 0:irrelevant, 1:so-so, 2:good, 3:excellent. Normalized Discounted Cumulative Gain ($NDCG$) is adopted to measure the visual relevance[3]. Giving a ranking list, the score $NDCG@n$ is defined as

$$NDCG@n = Z_n \Sigma_{i=1}^{n} \frac{2^{r(i)} - 1}{log(1+i)} \tag{14}$$

$r(i)$ is the score of the image in the $i^{th}$ rank order. $Z_n$ is the normalization factor to normalize the perfect rank to 1.

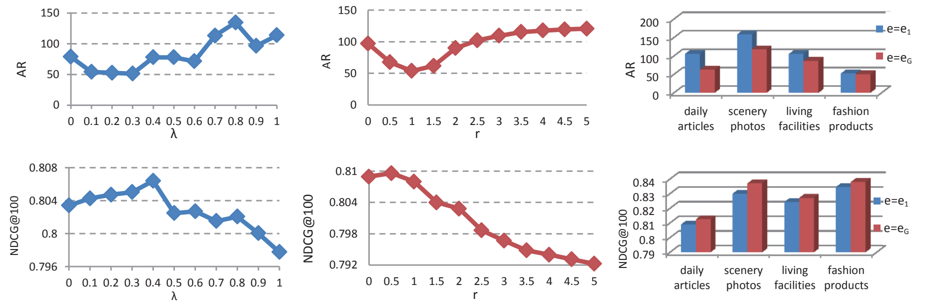


**Fig. 3.** Parameter settings for $\lambda$, $r$ and $e$ with $\alpha = 0.3$.Best performance is obtained with $\lambda = 0.4$, $r = 0.5$ and $e = e_G$.

### 4.3 Parameter Settings

In our approach, there are four parameters: $\lambda$ in Eq.(2), $r$ in Eq.(6), $\alpha$ in Eq.(1) and $e$ in Eq.(10). To study the effect of one parameter, we fix three other parameters as constants. Iteratively, we can find the optimal values for all the parameters to achieve the best performance.

From the Fig.3 we can find that our approach obtains the best performance when $\lambda = 0.4$, $r = 0.5$, $\alpha = 0.3$ and $e = e_G$. As the parameter representing the trade-off between the users' overlap and the images' overlap to determine group similarity, the value of $\lambda$ shows users are more likely to be interested in a group because of its images rather than users. The value of $r$ indicates that the importance of a group has small impact on visual relevance. In other words, an important group may also share some low-quality images. Besides, it can be observed that the algorithm with $e = e_G$ is significantly better than $e = e_1$ for all categories of queries. Thus, personalized vector $e_G$ can indeed improve the performance of our approach.

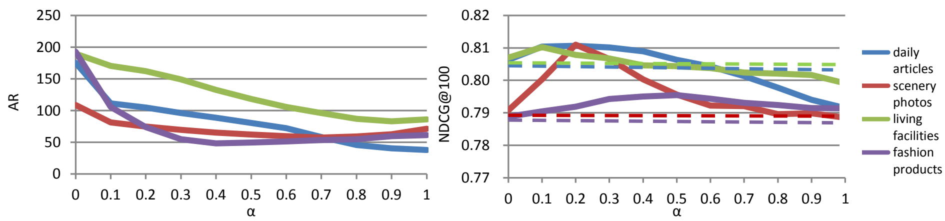


**Fig. 4.** Performance for different values of $\alpha$ with $\lambda = 0.4$, $r = 0.5$ and $e = e_G$. Best performance is obtained with $\alpha = 0.3$.

$\alpha$ is an important parameter to balance social factor and visual factor. We estimate the setting of $\alpha$ for each of the four categories. From the results in

Fig.4, we can observe following: (1)For any category, best $\alpha$ is not small. i.e., social factor is helpful in image search.(2)Measured by $AR$, larger $\alpha$ produces better performance. Therefore, social factor can improve the images' relevance to user interests.(3)The curve of $NDCG$ indicates that, as the weight of social factor growing after a critical point, more images with low visual relevance are ranked to the front. Based on these observation, $\alpha$ is determined to be 0.3 in our approach, which can guarantee a reasonable balance between social relevance and visual relevance.

### 4.4 Results and Performance

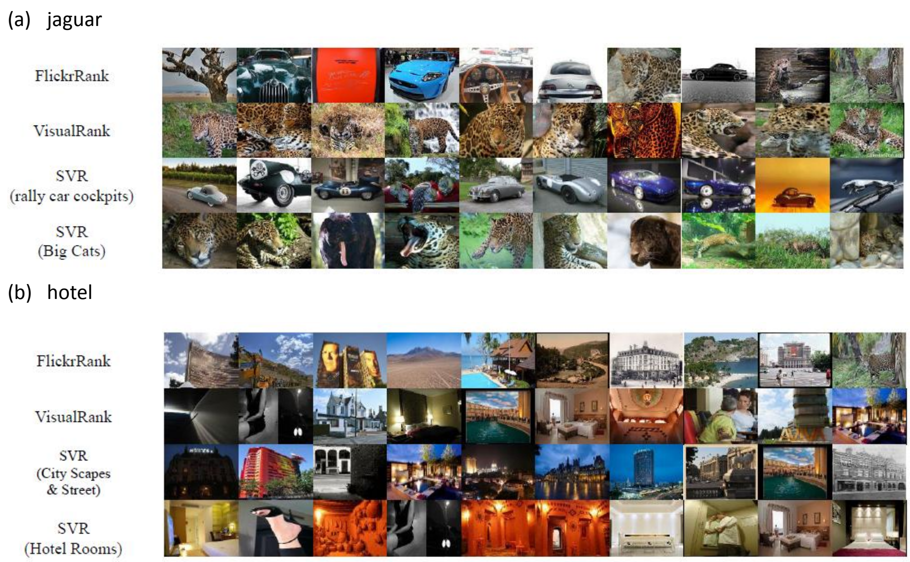


**Fig. 5.** Top-10 reranking results of our approach for two different groups compared to FlickrRank and VisualRank for two typical queries .

To prove the results of SVR can really reflect the user intentions, we select 2 queries "jaguar" and "hotel" to show cases of our results. For each query, we select 2 groups that we can obviously estimate the interests by their group names. Fig.5 shows the results. The content in the bracket after SVR is the group name. It can be observed that our approach really knows what the users want and the results are mostly of high quality. For the query "jaguar", which has obvious different concepts, SVR can find the images fit for the group names fairly well. In contrast, the top-10 results of VisualRank for "jaguar" are all about leopards.

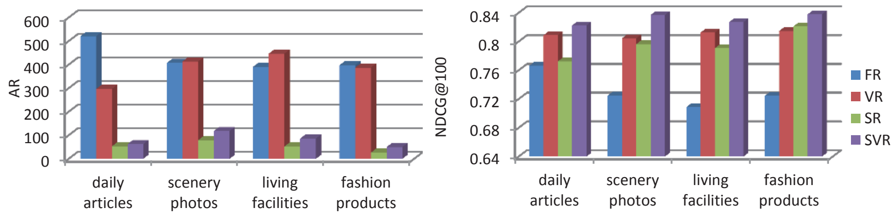


**Fig. 6.** The performance of our approach compared to other two methods FlickrRank and Visual Rank by the measurements $AR$ and $NDCG@100$ for four categories of queries.

For the quantitative evaluation of the performance, we compare our approach with other three ranking methods. Fig.6 shows the comparison results. It can be observed that our approach achieves the best performance in $NDCG$ and has great improvement in $AR$ compared to VR. Although $AR$ of SR is the best, $NDCG$ of SR is much worse than VR. Under the comprehensive consideration, our approach performs the best in these four ranking methods.

## 5 CONCLUSIONS

In this paper, we propose a novel framework of community-specific Social-Visual image Ranking for Web image search. We explore to combine the social factor and visual factor together based on image link graph to improve the performance of social relevance under the premise of visual relevance. Comprehensive experiment shows effectiveness of our approach. In that, it is significantly better than VisualRank and Flickr search engine in social relevance as well as visual relevance. Besides, the importance of both social factor and visual factor is discussed in details.

## 6 ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China, No. 60933013 and No. 61003097; National Program on Key Basic Research Project, No. 2011CB302206 and National Significant Science and Technology Projects of China under Grant, No. 2011ZX01042-001-002. Thanks for the support of Tsinghua-Tencent Associated Laboratory.

## References

1. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33(1C6):309 – 320, 2000.

2. W. H. Hsu, L. S. Kennedy, and S.-F. Chang. Video search reranking via information bottleneck principle. In *Proceedings of the 14th annual ACM international conference on Multimedia*, MULTIMEDIA '06, pages 35–44, New York, NY, USA, 2006. ACM.
3. K. Järvelin and J. Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '00, pages 41–48, New York, NY, USA, 2000. ACM.
4. Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(11):1877 –1890, nov. 2008.
5. R. I. Kondor and J. Lafferty. Diffusion kernels on graphs and other discrete structures. In *In Proceedings of the ICML*, pages 315–322, 2002.
6. M. Larson, C. Kofler, and A. Hanjalic. Reading between the tags to predict real-world size-class for visually depicted objects in images. In *Proceedings of the 19th ACM international conference on Multimedia*, MM '11, pages 273–282, New York, NY, USA, 2011. ACM.
7. X. Li, C. G. Snoek, and M. Worring. Learning tag relevance by neighbor voting for social image retrieval. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, MIR '08, pages 180–187, New York, NY, USA, 2008. ACM.
8. J. Liu, W. Lai, X.-S. Hua, Y. Huang, and S. Li. Video search re-ranking via multi-graph propagation. In *Proceedings of the 15th international conference on Multimedia*, MULTIMEDIA '07, pages 208–217, New York, NY, USA, 2007. ACM.
9. Y. Liu, T. Mei, and X.-S. Hua. Crowdreranking: exploring multiple search engines for visual search reranking. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 500–507, New York, NY, USA, 2009. ACM.
10. D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2161 – 2168, 2006.
11. L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous.
12. L. A. F. Park and K. Ramamohanarao. Mining web multi-resolution community-based popularity for information retrieval. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, CIKM '07, pages 545–554, New York, NY, USA, 2007. ACM.
13. J. Sang, J. Liu, and C. Xu. Exploiting user information for image tag refinement. In *Proceedings of the 19th ACM international conference on Multimedia*, MM '11, pages 1129–1132, New York, NY, USA, 2011. ACM.
14. M. Wang, B. Ni, X.-S. Hua, and T.-S. Chua. Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. *ACM Comput. Surv.*, 44(4):25:1–25:24, Sept. 2012.
15. M. Wang, K. Yang, X.-S. Hua, and H.-J. Zhang. Towards a relevant and diverse search of social images. *Multimedia, IEEE Transactions on*, 12(8):829 –842, dec. 2010.
16. W. Zhou, Q. Tian, Y. Lu, L. Yang, and H. Li. Latent visual context learning for web image applications. *Pattern Recognition*, 44(10C11):2263 – 2273, 2011. Semi-Supervised Learning for Visual Content Analysis and Understanding.