# TAG QUALITY IMPROVEMENT FOR SOCIAL IMAGES

*Dong Liu*[†] [*], *Meng Wang*[‡], *Linjun Yang*[‡], *Xian-Sheng Hua*[‡], *HongJiang Zhang*[§]

[†]School of Computer Sci.& Tec., Harbin Institute of Technology
[‡] Microsoft Research Asia
[§] Microsoft Advanced Technology Center

## ABSTRACT

Online social media sharing web sites like Flickr allow users to manually annotate images with tags, which can facilitate image search and organization. However, the tags provided by users are often imprecise and incomplete, which severely limits the application of tags to image search and browse. In this paper, we propose a scheme to improve poorly annotated tags associated with social images. Two properties are exploited and integrated in an unified optimization framework: (1) consistency between visual and semantic similarities, where the semantic similarity is estimated using tags; (2) compatibility of tags before and after improvement, since the initial user provided tags carry valuable information. An iterative bound optimization method is derived to solve the optimization problem. Experimental results on Flickr dataset show that the proposed method can significantly improve the quality of tags.

***Index Terms***— Tag, Improvement, Flickr, Social images

## 1. INTRODUCTION

Recent years have witnessed an explosion of community-contributed multimedia content available online. Such social media repositories (such as Flickr, Youtube and Zooomr) allow users to upload personal media data and annotate content with descriptive keywords called tags. With the rich tags as metadata, users can conveniently organize, search and index shared media content and this provides opportunities to make large-scale media retrieval work in practice.

We take Flickr [1] as an example to study the characteristics of social tagging. Filckr is one of the earliest and most popular social media sharing web sites and it has been intensively studied in recent years, especially on tagging characteristic [2, 3], tag recommendation [4, 5, 6], etc. A recent study in [7] reveals that users do annotate their photos with the motivation to make them better accessible to the general public. However, existing studies [8] show that the tags provided by Flickr users are highly noisy and there are only around 50% tags actually related to the image. Fig. 1 illustrates an exemplary image from Flickr and its tags. From the figure we can see that only "sky" and "cloud" correctly describe the content of the given image, and the other tags are imprecise(e.g., dog, girl, etc.) or subjective[1](e.g., family, city, etc.), Meanwhile, several other tags that can be useful, such as "tree" and "grass", have not been provided. The imprecise and incomplete tagging characteristics have significantly limited the access of social media. The imprecise tags will introduce false positives into user's search result and incomplete tags will make the actually related images inaccessible. Therefore, it would be advantageous if a dedicated approach can be developed to improve the tags associated with social images such that they can better describe the content of the images.


**Fig. 1**. An exemplar image from Flickr and its associated tags.

Some approaches [9, 10, 11] have been developed to refine annotation result for automatic image annotation algorithms. As a pioneering work, Jin et al. [9] have used WordNet to calculate the semantic correlation between annotation concepts and then highly correlated concepts are preserved and weakly correlated concepts are removed. However, this method has not considered the visual content clue, and it thus achieves only limited success. Several other works [10, 11] have leveraged both visual and textual clues to refine image annotation result, but these two clues are still only used to estimate the correlation between annotation concepts in order to perform belief propagation among concepts. Of course these methods can be directly applied in the tag quality improvement task. But they will not achieve satisfactory results since they have not explored the semantic consistency between visually similar images, i.e., the tags of similar images should be close[2] (empirical results in Section 4 will also demonstrate this fact).

In this work, we propose a tag quality improvement method based on the consistency of visual similarity and semantic similarity of images. Here the semantic similarity of two images is defined as the similarities of their tag sets (detailed formulation will be introduced in Section 2). In addition, we assume the compatibility of improvement process, i.e., the improved tags should not change too much from the initial tags. Thereby, we formulate an optimization framework that includes two terms to accomplish the task. We also propose an efficient algorithm to solve the optimization problem based on an iterative bound optimization method.

The contribution of this paper can be summarized as follows:

[*]This work was performed at Microsoft Research Asia.

[1]Subjective tags are those content unrelated tags that are not easily and consistently recognized by common knowledge.

[2]This is understandable for annotation refinement methods, since this consistency will be leveraged in annotation algorithms, i.e., the annotation results should already have this property. But this important property has to be utilized in tag quality improvemment process, since the semantic consistency will not be guaranteed for user provided tags.

- To the best of our knowledge, this is the first attempt to improve the unqualified tags of social images. In comparison with the existing image annotation refinement works [9, 10, 11], tag quality improvement will be more challenging due to the uncontrolled vocabulary of tags and the diversity of social images.
- We propose an optimization framework that simultaneously models the consistency of image visual and semantic similarities and the compatibility of tags before and after improvement. We also introduce an iterative upper bound optimization to solve it.

The rest of this paper is organized as follows. In Section 2, we provide the formulation of tag quality improvement. We then introduce its iterative optimization method in Section 3. In section 4, we provide empirical justification. Finally, we conclude this paper in Section 5.

## 2. OPTIMIZING TAG QUALITY

In this section, we introduce our tag quality improvement method. We firstly define some notations in Section 2.1 and the tag quality improvement scheme is discussed in detail in section 2.2.

### 2.1. Notations

Given a social image collection $D = \{x_1, x_2, \ldots, x_n\}$ with its associated all unique tags $\Omega = \{t_1, t_2, \ldots, t_m\}$. The initial tag membership for the whole image collection can be presented in a binary matrix $\hat{Y} \in \{0,1\}^{n\times m}$ whose element $\hat{Y}_{ij}$ indicates the membership of tag $t_j$ with respect to image $x_i$ (i.e., if $t_j$ is associated with image $x_i$, then $\hat{Y}_{ij} = 1$ and otherwise $\hat{Y}_{ij} = 0$). To represent the tag improvement result, we define another matrix $\mathbf{Y}$ whose element $Y_{ij} \geq 0$ denotes the confidence score of assigning tag $t_j$ to image $x_i$ and by $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{im})^\top$ the confidence scores of assigning each tag to the $i$-th image. Denote by $W$ a similarity matrix whose element $W_{ij}$ indicates the visual similarity between $x_i$ and $x_j$.

### 2.2. Formulation of Tag Quality Improvement

As previously mentioned, our tag quality improvement scheme is developed based on two assumptions, i.e., the consistency of visual and semantic similarities and the compatibility of the tags before and after improvement.

The visual similarity of images can be directly computed with Gaussian kernel function with a radius parameter $\sigma$, i.e.,

$$W_{ij} = exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}) \tag{1}$$

For semantic similarity, the most simple approach is to define it as the overlap between the tag sets, i.e., $\mathbf{y}_i^\top \mathbf{y}_j$. However, this approach actually treats all tags independently and cannot explore their correlation. For example, it will give zero similarity for two images that share no common tags, but in this case the images can still be very semantically similar if their tags are strongly correlated. To leverage the correlation of tags, we introduce matrix $S$ for the tag similarity, in which the element $S_{ij} \geq 0$ indicates the similarity between $t_i$ and $t_j$ (the computation of the similarity will be introduced in Section 4). Then we compute the semantic similarity of images by a weighted dot product, i.e., ${\mathbf{y}_i}^\top S \mathbf{y}_j = \sum_{k,l=1}^m Y_{ik} S_{kl} Y_{jl}$. Therefore, the consistency of visual and semantic similarities can be formulated as

$$\begin{aligned} &\mathbf{Y}^\star = \arg\min_{\mathbf{Y}} \textstyle\sum_{i,j=1}^n (W_{ij} - \sum_{k,l=1}^m Y_{ik} S_{kl} Y_{jl})^2 \\ &s.t. Y_{jl} \geq 0 \end{aligned} \tag{2}$$

A problem of the above equation is that the visual similarities and semantic similarities are at different scales (obviously the semantic similarities can be much larger than the visual similarities). Therefore, we introduce a scaling matrix $\Gamma$ to modulate the scales of visual similarities.(See Eq. 3)

We then consider the compatibility of tags before and after improvement. Here we use the minimization of $\sum_{j=1}^n \sum_{l=1}^m (Y_{jl} - \hat{Y}_{jl})^2 exp(\hat{Y}_{jl})$ to model this assumption. Note that we have adopted a weighting factor $exp(\hat{Y}_{jl})$. These weighting factors can let the compatibility term put more emphasis on the initially existed tags, i.e., the user-provided tags can be preserved with relatively high probability. Since the tag set of social images is usually very huge, this strategy can be useful and empirical study has demonstrate that this strategy can lead to better performance.

Considering both the consistency and compatibility terms, the optimization scheme can be formulated as

$$\begin{aligned} &[\mathbf{Y}^\star, \Gamma^\star] = \arg\min_{\mathbf{Y},\Gamma} \textstyle\sum_{i,j=1}^n (\Gamma_{ij} W_{ij} - \sum_{k,l=1}^m Y_{ik} S_{kl} Y_{jl})^2 \\ &+C \textstyle\sum_{j=1}^n \sum_{l=1}^m (Y_{jl} - \hat{Y}_{jl})^2 \exp(\hat{Y}_{jl}) \\ &s.t. Y_{jl} \geq 0, \Gamma_{ij} \geq 0, i,j = 1,2,\ldots,n, k,l = 1,2,\ldots,m. \end{aligned} \tag{3}$$

where $C$ is a weighting factor that modulates the effect of the two optimization terms.

## 3. SOLUTION OF THE OPTIMIZATION PROBLEM

We solve the optimization problem in Eq. 3 with an iterative bound optimization algorithm, which is analogous to [12]. We first deduce an upper-bound for the objective function and then approximate the optimal solution iteratively.

Firstly we will upper-bound the objective function in Eq. 3. We upper-bound $\sum_{i,j=1}^n (\Gamma_{ij} W_{ij} - \sum_{k,l=1}^m Y_{ik} S_{kl} Y_{jl})^2$ as follows

$$\begin{aligned} &\textstyle\sum_{i,j=1}^n \left(\Gamma_{ij} W_{ij} - \sum_{k,l=1}^m Y_{ik} S_{kl} Y_{jl}\right)^2 \\ &\leq \textstyle\sum_{i,j=1}^n \left(\sum_{k,l=1}^m \frac{\tilde{Y}_{ik} S_{kl} \tilde{Y}_{jl}}{[\tilde{Y} S \tilde{Y}^\top]_{ij}} (\Gamma_{ij} W_{ij} - [\tilde{Y} S \tilde{Y}^\top]_{ij} \frac{Y_{ik} S_{kl} Y_{jl}}{\tilde{Y}_{ik} S_{kl} \tilde{Y}_{jl}})^2\right) \\ &= \textstyle\sum_{i,j=1}^n \Big(\Gamma_{ij}^2 W_{ij}^2 + \sum_{k,l=1}^m (\frac{[\tilde{Y} S \tilde{Y}^\top]_{ij}}{\tilde{Y}_{ik} S_{kl} \tilde{Y}_{jl}} Y_{ik}^2 S_{kl}^2 Y_{jl}^2 \\ &-2\Gamma_{ij} W_{ij} Y_{ik} S_{kl} Y_{jl})\Big) \\ &\leq \textstyle\sum_{i,j=1}^n \Big(\Gamma_{ij}^2 W_{ij}^2 + \frac{1}{2} \sum_{k,l=1}^m [\tilde{Y} S \tilde{Y}^\top]_{ij} \tilde{Y}_{ik} S_{kl} \tilde{Y}_{jl} (\frac{Y_{ik}^4}{\tilde{Y}_{ik}^4} + \frac{Y_{jl}^4}{\tilde{Y}_{jl}^4}) \\ &-2 \textstyle\sum_{k,l=1}^m \Gamma_{ij} W_{ij} \tilde{Y}_{ik} S_{kl} \tilde{Y}_{jl} (1 + \log Y_{ik} + \log Y_{jl} \\ &- \log \tilde{Y}_{ik} - \log \tilde{Y}_{jl})\Big) \\ &= \textstyle\sum_{i,j=1}^n \Big(\Gamma_{ij}^2 W_{ij}^2 + \sum_{l=1}^m [\tilde{Y} S \tilde{Y}^\top]_{ij} [\tilde{Y} S]_{il} \frac{Y_{jl}^4}{\tilde{Y}_{jl}^3} \\ &-4 \textstyle\sum_{l=1}^m \Gamma_{ij} W_{ij} [\tilde{Y} S]_{il} \tilde{Y}_{jl} \log Y_{jl} - 2\Gamma_{ij} W_{ij} [\tilde{Y} S \tilde{Y}^\top]_{ij} \\ &+4 \textstyle\sum_{k=1}^m \Gamma_{ij} W_{ij} [S \tilde{Y}^\top]_{kj} \log \tilde{Y}_{ik}\Big) \end{aligned} \tag{4}$$

where $\tilde{Y}$ referrers to the matrix $\mathbf{Y}$ from the last iteration. We then upper bound the second term of Eq. 3 based on the concaveness of

logarithm function, i.e.,

$$\begin{aligned} & C\sum_{j=1}^{n}\sum_{l=1}^{m}(Y_{jl}-\hat{Y}_{jl})^2 exp(\hat{Y}_{jl}) \\ & = C\sum_{j=1}^{n}\sum_{l=1}^{m}(Y_{jl}^2-2\hat{Y}_{jl}Y_{jl}+\hat{Y}_{jl}^2) exp(\hat{Y}_{jl}) \\ & \leq C\sum_{j=1}^{n}\sum_{l=1}^{m}\left[Y_{jl}^2-2\hat{Y}_{jl}\tilde{Y}_{jl}(\log\frac{Y_{jl}}{\tilde{Y}_{jl}}+1)+\hat{Y}_{jl}^2\right] exp(\hat{Y}_{jl}) \end{aligned} \tag{5}$$

Finally, we combine two upper bounds in Eq. 4 and Eq. 5 as the final upper bound of Eq. 3. Taking the derivatives of the combined bounding function with respect to $Y_{jl}$ and $\Gamma_{ij}$ and set them to zero, we can obtain the following solution:

$$\begin{cases} Y_{jl} = \left[\frac{-C exp(\hat{Y}_{jl})\tilde{Y}_{jl}^3+\sqrt{M}}{4[\tilde{Y}S\tilde{Y}^\top\tilde{Y}S]_{jl}}\right]^{\frac{1}{2}} \\ \Gamma_{ij} = \frac{[\tilde{Y}S\tilde{Y}^\top]_{ij}+2G}{W_{ij}} \end{cases} \tag{6}$$

where $M = \left(C\exp\left(\hat{Y}_{jl}\right)\right)^2 + 8U_{jl}\tilde{Y}_{jl}^4(2[W\tilde{Y}S]_{jl} + C\hat{Y}_{jl}\exp\left(\hat{Y}_{jl}\right))$ with $U_{jl} = [\tilde{Y}S\tilde{Y}^\top\tilde{Y}S]_{jl}$ and $G = \sum_{l=1}^{m}[\tilde{Y}S]_{il}\tilde{Y}_{jl}\log Y_{jl} - \sum_{k=1}^{m}\tilde{Y}_{ik}[S\tilde{Y}^\top]_{kj} log\tilde{Y}_{ik}$.

With the obtained solution in Eq. 6, we apply an iterative process to get an approximate optimization solution. We firstly initialize $\mathbf{Y}$ and $\Gamma$ randomly, and then iteratively update the two matrices until convergence.

## 4. EXPERIMENTS

### 4.1. Experiment Setup

We conduct experiments with the data collected from Flickr. We select the ten most popular queries, including *cat*, *sky*, *mountain*, *automobile*, *water*, *flower*, *bird*, *tree*, *sunset* and *sea*, and use them as query keywords to perform tag-based search on Flickr. The search results are displayed using "ranking by interestingness" option. Then the top 1000 images for each query are collected together with their associated information including tags, user ID, etc. In this way, we obtain a social image dataset comprising of $10,000$ images and $38,335$ unique tags. However, many of the raw tags are misspelling and meaningless. Hence, we firstly adopt a pre-filtering process for these tags. Specifically, we match each tag with the entries in a Wikipedia thesaurus and only the tags that have a coordinate in Wikipedia are kept. In this way, 343 tags are kept for our tag refinement experiment. For each image, we extract 428-dimensional features, including 225-dimensional block-wise color moment generated from 5-by-5 partition of the image, 128-dimensional wavelet texture feature and 75-dimensional shape feature. The radius parameter $\sigma$ in Eq. 1 is set to the median value of all pair-wise Euclidean distances between images, and the parameter $C$ in Eq. 3 is empirically set to 100. In this work, we compute the similarity between tags $t_i$ and $t_j$ based on their co-occurrence analogous to Google similarity distance [13], i.e.,

$$S_{ij} = exp(-\frac{max(logf(t_i), logf(t_j)) - logf(t_i, t_j)}{logG - min(logf(t_i), logf(t_j))}) \tag{7}$$

where $f(t_i)$ and $f(t_j)$ are the numbers of images containing tag $t_i$ and tag $t_j$ on Flickr respectively and $f(t_i, t_j)$ is the number of images containing both $t_i$ and $t_j$ on Flickr (these numbers can be obtained by performing search by tag on Flickr website using the tags as keywords), and $G$ is the total number of images in Flickr. But it is worth noting that our method is flexible, and the similarities can be computed through other approaches, such as using WordNet or Flickr distance [14]. On the other hand, the tag correlation can be more precisely estimated if we further consider each tag's relevance with respect to its associated image, such as in [15, 16].

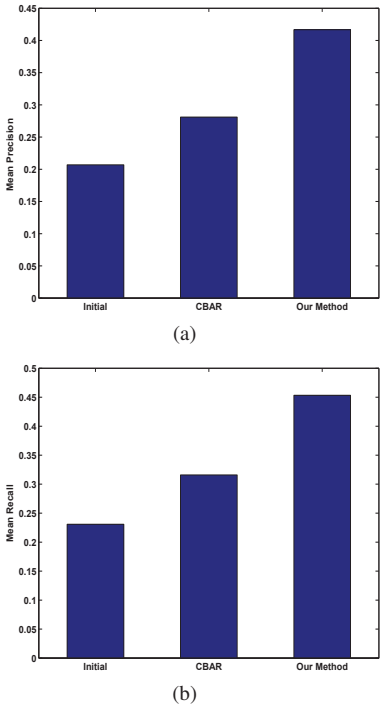



**Fig. 2**. Comparison of original tags, the results improved using CBAR and using our method in term of mean precision and recall.

To evaluate the performance of our tag quality improvement scheme, we calculate precision and recall of the tag set of each image and then average them as final evaluation measure. Each tag is judged by human labelers to decide whether it is related to the image. However, manually labeling all images and the tags before and after improvement is too labor-intensive, and thus we only randomly select 500 images as evaluation set.

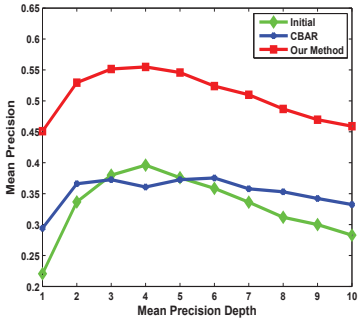
### 4.2. Experiment Results

We compare the following three results:

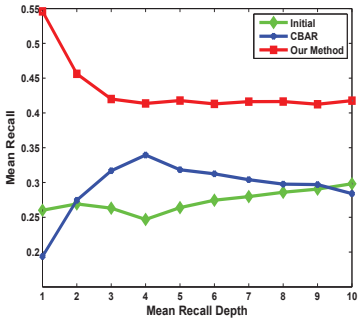
- The original tags, i.e., the baseline.
- The tags produced by a content based annotation refinement approach proposed in [10] ("CBAR" for short).
- The tags produced by our method.

Fig. 2 shows the mean precision and recall of the results. From the figure we can see CBAR and our method have achieved improvements in comparison with the baseline. But the improvement of CBAR is limited. As analyzed in Section 1, this is due to the fact that it has not sufficiently explored visual information. Our method performs much better in both precision and recall. We also illustrate the results at different depths in Fig. 3. Specifically, we only keep $k$ tags for each image(for original tags, we used their order in Flickr, and for refined results we order them according to the finally obtained confidence scores) and then evaluate the mean precision and

recall. From the figure we can clearly see that our method consistently outperforms the other two methods. Several exemplary images and their tags before and after improvement are illustrated in Fig. 4.




(a)




(b)

**Fig. 3**. Comparison of different results with varied depths.

## 5. CONCLUSION

As an initial effort towards improving tagging quality for social images, a tag quality improvement scheme has been proposed. The proposed approach simultaneously models the consistency of visual and semantic similarities as well as compatibility of tags before and after improvement in an integrated optimization framework. An effective iterative process is then introduced to solve the optimization problem. Encouraging results are reported, which demonstrate the effectiveness of our tag quality improvement approach. In the future, we plan to explore tag quality improvement problem in a more general scenario, including tag categorization, enrichment and ranking, aiming to build better lexical indexing for social images.

## 6. REFERENCES


[1] Flickr. *http://www.flickr.com.*

[2] C. Marlow, M. Naaman, D. Boyd and M. Davis. HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, Toread. In *Proceedings of the 17th Conference on Hypertext and Hypermedia*, 2006.

[3] T. Rattenbury, N. Good and M. Naaman. Towards Automatic Extraction of Event and Place Semantics from Flickr Tags. In *Proceeding of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2007.

[4] B. Sigurbjörnsson, R. V. Zwol. Flickr Tag Recommendation based on Collective Knowledge. In *Proceeding of ACM International World Wide Web Conference*, 2008.

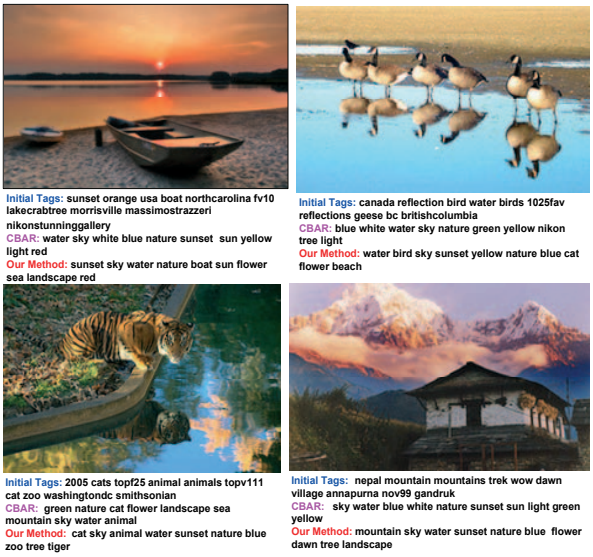




**Fig. 4**. Several exemplary images and their tags before and after improvement. We can see that the tags produced using our method are obviously much more precise.


[5] N. Garg, I. Weber. Personalized, Interactive Tag Recommendation for Flickr. In *Proceeding of ACM International Conference on Recommender Systems*, 2008.

[6] K. Weinberger, M. Slaney and R. V. Zwol. Resolving Tag Ambiguity. In *Proceeding of 15th ACM International Conference on Multimedia*, 2008.

[7] M. Ames and M. Naaman. Why We Tag: Motivations for Annotation in Mobile and Online Media. In *proceeding of the SIGCHI Conference on Human Factors in Computing System*, 2007.

[8] L. S. Kennedy, S. F. Chang, I. V. Kozintsev. To Search or To Label? Precdicting the Performance of Search-Based Automatic Image Classifiers. In *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, 2006.

[9] Y. Jin, L. Khan, L. Wang, and M. Awad. Image Annotation by Combining Multiple Evidence & WordNet. In *Proceeding of 12th ACM International Conference on Multimedia*, 2005.

[10] C. Wang, F. Jing, L. Zhang and H. J. Zhang. Content-Based Image Annotation Refinement. In *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[11] C. Wang, F. Jing, L. Zhang and H. J. Zhang. Image Annotation Refinement using Random Walk with Restarts. In *Proceeding of the 14th ACM International Conference on Multimedia*, 2006.

[12] Y. Liu, R. Jin and L. Yang. Semi-supervised Multi-label Learning by Constrained Non-Negative Matrix Factorization. In *Proceeding of the 21st National Conference on Artificial Intelligence*, 2006.

[13] R. Cilibrasi, P. M. B. Vitanyi. The Google Similarity Distance. In *IEEE Transactions on Knowledge and Data Engineering*, 2007.

[14] L. Wu, X. S. Hua, N. H. Yu, W. Y. Ma and S. P. Li. Flickr Distance. In *Proceeding of 15th ACM International Conference on Multimedia*, 2008.

[15] X. R. Li, C. G. M. Snoek, and M. Worring. Learning Tag Relevance by Neighbor Voting for Social Image Retrieval. In *Proceedings of the ACM International Conference on Multimedia Information Retrieval*, 2008.

[16] D. Liu, X. S. Hua, L. J. Yang, M. Wang and H. J. Zhang. Tag Ranking. In *Proceeding of ACM International World Wide Web Conference*, 2009.