

# Personalized Annotation for Mobile Photos Based on User's Social Circle

Yanhui Hong, Tiandi Chen, Kang Zhang, and Lifeng Sun<sup>(✉)</sup>

Tsinghua National Laboratory for Information Science and Technology,  
Department of Computer Science and Technology,  
Tsinghua University, Beijing, China  
hongyh13@mails.tsinghua.edu.cn, dalianctd@163.com, zk54188@gmail.com,  
sunlf@tsinghua.edu.cn

**Abstract.** For mobile photos annotation, users are more interested in the context information behind the photos. The user's social circle can provide valuable information for it. However, the accompanying textual information of social network is sparse and ambiguous in nature. In this paper, we propose a personalized annotation framework for mobile photos leveraging the user's social circle. To address the unreliability problem of social network, we present an algorithm to generate reliable tags for social photos before assigning tags to the user's unlabeled photos. In the tag generation stage, a multi-modality hierarchical clustering algorithm is performed to detect social events. Besides, we use "Album" instead of individual photo as the basic unit for clustering. Finally, we employ a weighted nearest neighbor model for label propagation. We evaluate our framework on a large-scale, real-world dataset from Renren, the largest Facebook-like social network in China. Our evaluation results show promising results of our proposed framework.

**Keywords:** Personalized annotation · Social network · Event detection

## 1 Introduction

With the prevalence of mobile devices, people can take photos at anytime and anywhere. And owing to the screen and memory limitations of mobile devices, people are used to transmitting and storing their taken photos via third-party cloud services, such as iCloud. How to organize and manage these personal photos has become a much more pressing issue for users due to the large data set size.

Image annotation is an effective and promising technique to solve this problem. However, full manual image annotation is labor-intensive and time-consuming. Thus, automatic image annotation is crucial and has received a lot of research interests. Different from general-purpose photo annotation algorithms [13, 23], which try to assign general visual labels to the images, such as "cats" or "boys", mobile photos annotation is a highly personal and user-centric task, for example users may more concern about "My brother Jack" instead of "a boy", "my graduation trip to

Hawaii” instead of “beach”. The question is then how to get the context information behind the photos? Undoubtedly, the user's social circle is an important source since there are millions of users upload and share their personal photos in the social media platforms, such as Flickr, Facebook and Twitter. And there is a significant overlap in their real world activities as the participants in the network are often family, friends or co-workers. For a photo to be labeled in the user's mobile device, it is probably that there are social friends who participate in the same activity and upload the related photos to the social network. The community-contributed photos with their associated social information will be of tremendous value for personalized annotation.

In this paper, we aim to provide personalized annotation for mobile photos based on social network. However, it is a very challenging task because: (1) The accompanying tags of social images are noisy and heterogeneous in nature since they are annotated by different users, and far from uniform in quality and might often be misleading or ambiguous; (2) Given the diversity of social network, the tags are more personal and the number of tags to be modeled in this situation is larger than that of predefined labels in typical annotation problem, which renders most of the traditionally approaches undesirable for our scenario.

Therefore, we would like to propose a new framework to address the mentioned problems. The proposed framework is based on two observations:

1. Event is one of the most important elements of people's life and memories. And most of the personal photos are taken during specific events such as birthday party, family trip, sport meeting etc. The same event should share the same event attribution labels, such as what, where, when and who.
2. Besides, event is highly time dependent. And unlike the images from web searching engine or commercial image banks, the photos in user's social circle are related to each other and the photos uploaded by the same user at the same time are most likely belong to the same event.

Base on observation 1, we can come to conclusion that although the associated social information of a photo in social network may be missing or ambiguous, combing all the social information of photos belong to the same event will provide more reliable labels. So we detect social events at first to generate reliable event tags for social photos. What's more, considering the sparsity and unreliability of individual social photo and being inspired by observation 2, we use “Album” as the basic unit for clustering in social event detection stage innovatively - it can be the album structure on some social platforms or a manually defined way that the photos uploaded by the same user at the same time.

Based on the above analyses, we propose a personalized annotation framework based on the user's social circle. Our system can be divided into two main parts, label generation and image annotation. In the label generation stage, the events in the user's social circle are detected by a novel multi-modality hierarchical clustering algorithm at first. Different from the previous works, we exploit the intrinsic properties of social network and events for event detection. In our multi-modality hierarchical clustering algorithm, a temporal-based

clustering algorithm is performed after separating photos into albums. And, a multi-modality agglomerative hierarchical clustering algorithm is employed in each temporal cluster result respectively. Then the representative labels for each event are extracted from the textual in the same cluster. After doing it, each photo in the user’s social circle will be associated with some reliable event labels and its initial tags given by the uploader. In the image annotation stage, personalized labels for the mobile photos are generated by a weighted K-nearest neighbor model similar to [6]. We improve the model by using both visual and date information to get the neighbors.

The contributions of this paper are manifold:

1. To tackle the unreliability issue of tags in social network, we exploit the characteristic of personal social circle and detect social events at first to generate reliable event tags.
2. We use “Album” as the basic unit for clustering and event detection, which not only eases the problem of large scale clustering, but also addresses the problem caused by the unreliability and sparsity of individual photo.
3. We propose a novel hierarchical clustering algorithm exploiting multi cues including the content, time, textual and social behavior information for social event detection.

The rest of the paper is organized as follows: In Sect. 2, the related works is introduced. In Sect. 3, we present our personalized annotation system, detail with the label generation and propagation. Data set analysis and experimental results are shown in Sect. 4. In Sect. 5, we conclude the paper and discuss future work.

## 2 Related Work

In this section, we briefly review some existing literature related to our work.

**Content-Based Annotation.** In recent years, many content-based annotation algorithms have been proposed and dramatically advanced this field [4]. They can be categorized into three main groups: (1) generative models [15], which try to estimate the joint probabilities between image visual features and labels; (2) discriminative models [7, 21], which regard image annotation as a classification problem and consider each pre-defined label as an independent class; (3) graph-learning models [10, 19], which use label diffusion over a similarity graph of labeled and unlabeled images. However, all of these content-based methods suffer from the well-known semantic gap.

**Social Event Detection.** As mentioned above, social event detection is performed to generate event tags for social photos in our label generation stage. Although there are some work focus on event detection from photos social meta-data, most part of the methods regard it as an event classification or recognition problem in which the event ontology and number are pre-defined [2]. It is not suitable for our situation where there may be hundreds or thousands events in the user’s social circle and the events are unknown before detection.

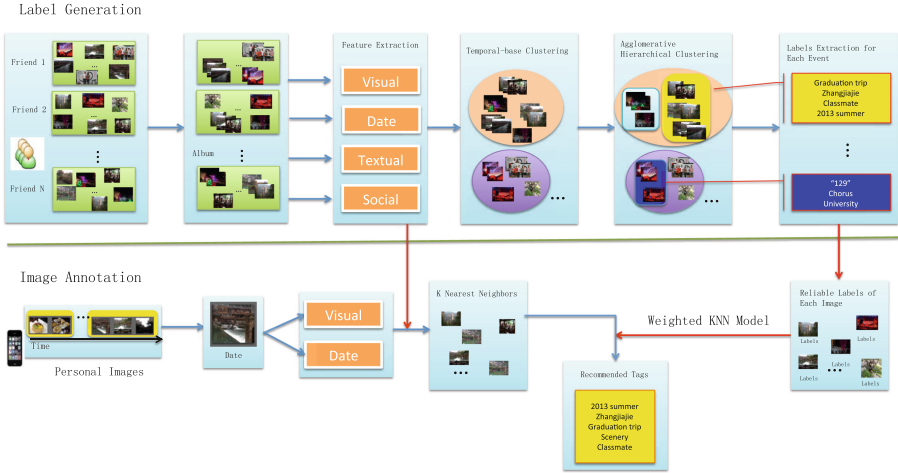
Recently, there are some works try to detect social event by clustering. In [1], the authors employ both ensemble and classification-based similarity learning techniques in conjunction with an incremental clustering algorithm to solve this problem, which is naive and only the textual, location and date information is used. In [16], the authors use pairwise similarities to predict a “same cluster” relationship. However, a known clusters from the same domain is required to adjust the weights and finally a K-mean or spectral cluster is performed, which requires a priori knowledge of the cluster number.

**Personalized Annotation.** To provide personalized annotation, the rich contextual information has been investigated. Some works leverage other mobile applications such as weather API, personal calendar or email context to get personalized labels [5]. For example, the calendar entry “Bob’s birthday party on July 12, 2013” provides strong complementary information. In [3], the GPS location, compass direction and image visual are employed to find potential point of interest (POI) in a given area by clustering. A cross-entropy based learning algorithm to personalize a generic annotation model is proposed in [9]. Whereas in [12], a personalized tag recommendation system is proposed, which takes users’ characteristics and tagging habits into consideration and gets the tag list by tags voting. [11] proposes a unified framework using subspace learning method to suggest personalized and geo-specific tags dynamically. The intuitive efforts have obtained a certain success in some extend, but the information they can exploit is limited and most of them are relying on the user’s tagging history or geographical information.

**Social Annotation.** Leveraging social data for photos annotation has attracted significant attention recently. In [22], the authors propose a graph-learning based personalized annotation framework leveraging the friends social network photos as training dataset. There are also some works trying to exploit the social behavioral information, such as comments and likes. In [14], the authors employ social metadata (common galleries, locations, uploaders) to extend the SVM model to include relational features, with the intuition that images sharing common properties are likely to share labels. In [8], a common-interest model is presented, which studies the common interests between pairwise users from the social diffusion records of sharing content. However, they always ignore the sparsity and unreliability of social metadata and don’t exploit the wealth of events’ intrinsic properties, which is the highlight of our framework.

### 3 Proposed Framework

As mentioned above, for mobile photos annotation, users are more interested in the context information behind the photos, which makes it different from the previous content-based image annotation works. On the other hand, the user’s social circle can provide valuable information. Intuitively, similar images are more likely to have the same labels. Starting from this intuition, we propagate the tags in the user’s social circle to personal unlabeled images based on image similarity. However, as is known to all, the textual information contributed by



**Fig. 1.** Overview of our system

common users in the social network is sparse, ambiguous and unreliable. So, before label propagation, we should tackle this issue by generating reliable tags for social photos. Based on the observations and analyses in Sect. 1, we detect social events to generate high confidence event labels at first. For social event detection, we develop a multi-modality hierarchical clustering algorithm exploiting the intrinsic properties of social network and using “Album” as the basic clustering unit. Hence, our system can be divided into two main parts: label generation and image annotation. Figure 1 provides an overview of our system.

To summer, the process of our system is:

## 1. Label Generation

- 1.1 Separate all photos in the user’s social circle into albums.
- 1.2 Extract visual, date, textual and social features, and learning multi-feature similarity metrics among albums.
- 1.3 Hierarchical clustering based on albums, including density-based clustering according to the photos’ taken date and multi-modality hierarchical agglomerative clustering in each temporal cluster result.
- 1.4 Generate representative event labels in each cluster for all photos.

## 2. Image Annotation

- 2.1 Train the weighted nearest neighbor model with discriminative metric learning from the user’s social circle data as well as generated tags.
- 2.2 Extract the visual and date feature of the given personal photos.
- 2.3 Get k-nearest-neighbors of the given photo according to its visual and taken date information.
- 2.4 Predict tags from the trained weighted k-nearest-neighbors neighbors by label propagation.

In the following, we provide more details on each of these components.

### 3.1 Label Generation

**Album.** The current social event detection algorithms suffer from two problems: the scalability problem and the unreliability of individual photo. How to address the two problems is a big challenge in our work.

Different from the images from web searching engine or commercial image banks, the photos in user’s social circle are related to each other. And some social platforms even provide “Album” structure for users to organize their photos, such as “Renren”. Experientially, the photos that a user uploads to the same album usually belong to the same event. What’s more, even without the “Album” structure in social platforms, we can find that the photos uploaded by the same user at the same time are often closely correlated to each other, and likely to be taken at the same event. So, we introduce “Album” concept, it defined as a photo set, in which all the photos share the same event attribution and uploaded by the same user. Namely, the photos in the same “Album” belong to the same event, and an event may consist of one or more “Albums”. Our experiments in Sect. 4 also indicate it.

So, instead of processing photo-by-photo, we use “Album” as basic unit for clustering to address the problem caused by the unreliability of individual photo and ease the pressure due to the large database.

In our paper, albums are generated by the following way: (1) If the social platforms have “Album” structure, then the photos in an album of social platform are regarded as belonging to the same album; (2) If the social platforms do not have “Album” structure, for each user in the social network, we get all the user’s upload photos and sort them according to their upload dates. Then if the upload dates of two adjoining photos are within an hour, they are regarded as in the same album, otherwise they belong to different albums and we create a new album for the second photo. There are some works using the similar concept to us. However there are some fundamental differences between our “Album” concept and them. For example, in [20] the authors use Flickr Groups to provide more accurate annotation. They train specific annotation models for different Flickr groups (like Rome or Wedding), and chose the trained most appropriate Flickr group to generate labels for a given batch of images. They only use the common style of Flickr Groups, while we stress the concept of events.

**Multi-modality Feature Extraction.** As a distinctive characteristic, social networks include a variety of context features [1], which will help our clustering task. It is noteworthy that as stated above, we utilize “Album” as the base clustering unit, the similarity measures should be defined between two albums instead of two photos. In our work, the following features and corresponding similarity measures are used:

- **Visual Feature:** We use the 4096-dimensional visual feature vector extracted by Convolutional neural networks [7] to represent the image visual information, which has been widely used for different recognition problems. The visual similarity of two albums is defined as follow:

$$S_v(A, B) = \frac{1}{|A|} \sum_{i=1}^{|A|} \max_j \{v(a_i, b_j)\} \quad (1)$$

where  $A, B$  are two albums and  $a_i, b_j$  are the photos in  $A$  and  $B$ ;  $v(a_i, b_j)$  is the visual similarity of photo  $a_j$  and  $b_j$  defined as the cosine distance of the visual feature vectors of two images. Noted that  $S_v(A, B)$  may be not equal to  $S_v(B, A)$ , so we use the average of  $S_v(A, B)$  and  $S_v(B, A)$  as the visual similarity score of albums  $A$  and  $B$ .

- **Date Feature:** We represent date as the number of minutes elapsed since the Unix epoch. Let  $t_a, t_b$  be the date value of image  $a$  and  $b$ , their similarity is defined as:  $s_t = 1 - |t_a - t_b|/T$ , where  $T$  equals  $365 \times 24 \times 60$ , namely  $T$  is the number of minutes in a year. If  $s_t < \epsilon$ , then  $s_t = \epsilon$  (We use  $\epsilon$  to avoid non-positive similarity score, and in practice we set  $\epsilon = 10^{-6}$ ). The date similarity of two albums is defined as follow:

$$S_t(A, B) = \max(1 - \max(\frac{D_B^{min} - D_A^{max}}{T}, \frac{D_A^{min} - D_B^{max}}{T}), 0, 0) \quad (2)$$

where  $D_A^{min} = \min_i \{t_{a_i}\}$ ,  $D_A^{max} = \max_i \{t_{a_i}\}$ ,  $D_B^{min} = \min_i \{t_{b_i}\}$ ,  $D_B^{max} = \max_i \{t_{b_i}\}$ . Intuitively speaking,  $S_t(A, B)$  is the time span of album  $A$  and album  $B$ .

- **Textual Feature:** There are various textual features accompanying social photos, such as tag, title and description. They can be transformed into words by extracting nouns using natural language processing techniques. We defined a weighted similarity metric for different texts as they are unreliable at different level, for example title may provide strong complementary information than description. The weighted Jaccard similarity coefficient is employed to measure textual similarity. The textual feature similarity of two albums is defined as follow:

$$S_w(A, B) = w_{tag} * J_{tag} + w_{title} * J_{title} + w_{desc} * J_{desc} \quad (3)$$

where  $J_{tag}$  is the Jaccard similarity of tags in album  $A$  and album  $B$ , defined as follow:

$$J_{tag} = \frac{|Tag_A \cap Tag_B|}{|Tag_A \cup Tag_B|} \quad (4)$$

$J_{title}$  and  $J_{desc}$  are defined the same as  $J_{tag}$ . And  $w_{tag}$ ,  $w_{title}$  and  $w_{desc}$  are the weights and  $w_{tag} + w_{title} + w_{desc} = 1$ .

- **Social Feature:** We estimate social similarity according to multiple social factors such as friend relationship, comments, favorite images and share behavior. Let  $U_a, U_b$  be the owners of album  $A$  and  $B$ . The social similarity of two albums is defined as follow:
  - If  $U_a$  and  $U_b$  are social friends, then their friend similarity is 1, otherwise is 0;
  - If  $U_a$  comments photos in album  $B$  or  $U_b$  comments photos in album  $A$ , then their comment similarity is 1, otherwise is 0;

- If  $U_a$  favorites photos in album  $B$  or  $U_b$  favorites photos in album  $A$ , then their favorite similarity is 1, otherwise is 0;
- If  $U_a$  shares photos in album  $B$  or  $U_b$  shares photos in album  $A$ , then their share similarity is 1, otherwise is 0;

Having defined all these feature representation and corresponding similarity metrics, we combine all the features using a weighted similarity consensus function.

**Hierarchical Clustering and Event Representation.** As a key contribution, we propose a novel hierarchical clustering algorithm to detect social event.

For our scenario, the clustering algorithms should be scalable and not require a priori knowledge of the cluster number. So the traditional clustering algorithms required cluster numbers, such as K-means and spectral clustering, are not suitable in our situation.

Note that the events in our case are always small and there may be hundreds or thousands events in the user’s social circle and many of them are hosted by few users. So the agglomerative hierarchical clustering is preferable for our clustering task, which is performed based on album similarity.

Considering that events are always time depended and do not last long, we employ a temporal-based clustering at first and perform agglomerative hierarchical clustering algorithm in each date cluster result. By doing it, the data scale is reduced and clustering performance is improved due to the less noise.

For temporal-based clustering, we exploit a density-based algorithm base on [18]. In this stage, we use photo instead of “Album” as the basic unit. The local density  $\rho_i$  is calculated by a Gaussian kernel based on date similarity for each photo. Then the minimum date distance between the photo  $i$  and any other photos with higher density  $\delta_i$  is calculated. In our work, if the value of  $\rho$  or  $\delta$  is larger than a pre-defined threshold, we then create a new cluster for it.

We exploit agglomerative hierarchical clustering on each temporal-based cluster result, and merge them. For agglomerative hierarchical clustering, we combine all the feature similarities by a weighted function as a final similarity score between two albums. The process is as follow:

1. Each album is regarded as a separated cluster at first.
2. The two clusters with smallest distance are selected and merged into one cluster.
3. Calculate the similarity score between the new merged cluster and other clusters.
4. Repeat Step 2 and 3 until the smallest distance larger than a pre-defined threshold.

Then the representative labels for each event are extracted from the textual in the same cluster. And all the photos in the same cluster share the labels.



### 3.2 Label Propagation

As we mentioned, there are many approaches for annotation from labeled images. However, the discriminative models which should learn classifier for each label are unsuitable for our problem, since the labels of personalized annotation are heterogeneous and highly user-centric. Besides, generative models require a strong correlation relationship between images and tags, while the labels in our scenario are subjective. Intuitively, images have similar properties are likely to share labels. Graph-based label propagation algorithm is preferable for our annotation task, as it does not take the labels’ inherent meaning into consideration. However, full graph method is time consuming due to the large data set. To tackle this issue, K-nearest-neighbor like methods have been introduced, which predict tags taking a combination of the tag absence/presence among neighbors.

So, we employ a weighted K-nearest-neighbor model similar to [6] to predict tags for given personal photos. Since both visual and date feature may provide complementary information for annotation, we get the K nearest neighbors based on both the visual and date similarity.

## 4 Experiment

### 4.1 DataSet

We employ both the public available dataset ReSEED [17] and real-world dataset for our experiments.

The ReSEED dataset consists of pictures collected from Flickr and the corresponding metadata such as user information, upload and capture time, geographic information, tags, title and description. And all the pictures are assigned to individual social events. We use a subset of the dataset with a capture time between January 1, 2012 and December 31, 2012, yielding a dataset of 15577 pictures assigned to 714 events in total. The dataset is employed to verify the assumption about “Album”.

In this paper, we annotate a user’s mobile photos exploiting his (or her) personal social circle, which is unavailable in public datasets. To evaluate our system, we crawl images together with their context from the user’s social circle (Renren) from January 1, 2013 to December 31, 2013 and manually tag the event of each photo for evaluating our clustering algorithm. As a result, we construct a training data based on the user’s social circle and give annotation for the user’s personal images. Table 1 provides more details regarding the Renren dataset used in our experiments.

**Table 1.** Statistics of our real-world dataset

#Photos	#Friends	#Events	#Albums	#Min photos of a friend	#Max photos of a friend	#Min photos of an event	#Max photos of an event
33879	361	1174	1149	1	2663	1	1145

**Table 2.** The performance of Album

Dataset	Purity	Precision	Recall	F1
ReSEED	<b>0.9962</b>	<b>0.9992</b>	0.3986	0.5698
Renren	<b>0.9995</b>	<b>0.9998</b>	0.6241	0.7685

**Table 3.** Clustering performance comparison in terms of NMI and F1

Methods	NMI	F1
Incremental clustering [1]	0.7973	0.1983
Our algorithm	<b>0.9384</b>	<b>0.6381</b>

## 4.2 Evaluation of Album

The key hypothesis of this paper is that the photos belong to the same album are taken in the same event. Since the albums are generated by two different approaches as stated in Sect. 3.1, we analyze two datasets representing the two cases respectively.

For social media platforms with “Album” structure, we use the real-word dataset crawled from Renren. For social media platforms without “Album” structure, we use the ReSEED dataset from Flickr stated in Sect. 4.1. For the ReSEED dataset, we separate all the photos into albums according their taken time and get 1272 albums finally. To verify the proposed assumption and evaluate the performance of our album generation approach, we regard each album as a cluster and measure its performance using Precision and Purity. Table 2 shows the result. We can observe that the purity and precision scores are nearly 100% in both the two datasets, which demonstrates our assumption. Besides, note that album is not equivalent to event, an event can consist of many albums, so the recall and F1 scores are low.

## 4.3 Evaluation of Tag Generation

To demonstrate the advantages of our proposed hierarchical clustering algorithm, we use NMI and F1 to measure the performance comparing with a single-pass incremental clustering algorithm used in [1]. Table 3 shows the results. As it indicates, our proposed hierarchical clustering algorithm is much better than the baseline for both NMI and F1 score.

## 4.4 Evaluation of Personalized Annotation

In this section, we present a experimental comparison between the performance of content-based image annotation system [7], employing original unreliable social accompanying tags directly on the KNN model and the performance of our proposed framework, which generates reliable tags by social event detection before performing the KNN model.

**Table 4.** Comparison of Recall, Precision and F1

Methods	Recall	Precision	F1
Conted-based [7]	0.0363	0.0205	0.0262
Without tags generation	0.8252	0.4615	0.5919
Our proposed algorithm	<b>0.8752</b>	<b>0.5026</b>	<b>0.6385</b>

In our experiments, we combine all the top 5 ranked labels generated by the three methods, and allow users to select their preferred labels. Recall, Precision and F1 are adopted to measure the performance. Table 4 shows the result. We can observe that our proposed framework obtains the best performance.

## 5 Conclusion

In this paper, we proposed a personalized annotation framework for mobile photos leveraging the user’s social circle. To address the issue caused by the sparsity and unreliability of social photo tags, we generated reliable tags by detecting social events at first. An multi-modality hierarchical clustering algorithm using “Album” as the basic unit was proposed to detect social event by exploiting all the text, date, social behavior and visual features. By analyzing the characteristic of our scenario, a weighted KNN model was exploited to propagate the generated tags of social photos to the user’s unlabeled photos. Experimental results show our system is effective. In the future work, we will use the additional information in personal photos as feedback to refine the tag generation stage.

**Acknowledgments.** This work was part-funded by 973 Program under Grant No. 2011CB302206, National Natural Science Foundation of China under Grant No. 61272231, 61472204, Beijing Key Laboratory of Networked Multimedia.

## References

1. Becker, H., Naaman, M., Gravano, L.: Learning similarity metrics for event identification in social media. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 291–300. ACM (2010)
2. Cao, L., Luo, J., Kautz, H., Huang, T.S.: Image annotation within the context of personal photo collections using hierarchical event and scene models. *IEEE Trans. Multimedia* **11**(2), 208–219 (2009)
3. Cheng, A.J., Lin, F.E., Kuo, Y.H., Hsu, W.H.: Gps, compass, or camera? Investigating effective mobile sensors for automatic search-based image annotation. In: Proceedings of the International Conference on Multimedia, pp. 815–818. ACM (2010)
4. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv. (CSUR)* **40**(2), 5 (2008)
5. Gallagher, A.C., Neustaedter, C.G., Cao, L., Luo, J., Chen, T.: Image annotation using personal calendars as context. In: Proceedings of the 16th ACM International Conference on Multimedia, pp. 681–684. ACM (2008)

6. Guillaumin, M., Mensink, T., Verbeek, J., Schmid, C.: Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 309–316. IEEE (2009)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
8. Lei, C., Liu, D.: Image annotation via social diffusion analysis with common interests. In: 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 1–6. IEEE (2014)
9. Li, X., Gavves, E., Snoek, C.G., Worring, M., Smeulders, A.W.: Personalizing automated image annotation using cross-entropy. In: Proceedings of the 19th ACM International Conference on Multimedia, pp. 233–242. ACM (2011)
10. Liu, J., Li, M., Liu, Q., Lu, H., Ma, S.: Image annotation via graph learning. *Pattern Recogn.* **42**(2), 218–228 (2009)
11. Liu, J., Li, Z., Tang, J., Jiang, Y., Lu, H.: Personalized geo-specific tag recommendation for photos on social websites. *IEEE Trans. Multimedia* **16**(3), 588–600 (2014)
12. Liu, X., Qian, X., Lu, D., Hou, X., Wang, L.: Personalized tag recommendation for flickr users. In: 2014 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2014)
13. Makadia, A., Pavlovic, V., Kumar, S.: A new baseline for image annotation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 316–329. Springer, Heidelberg (2008)
14. McAuley, J., Leskovec, J.: Image labeling on a network: using social-network metadata for image classification. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part IV. LNCS, vol. 7575, pp. 828–841. Springer, Heidelberg (2012)
15. Niu, Z., Hua, G., Gao, X., Tian, Q.: Semi-supervised relational topic model for weakly annotated image recognition in social media. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4233–4240. IEEE (2014)
16. Petkos, G., Papadopoulos, S., Kompatsiaris, Y.: Social event detection using multimodal clustering and integrating supervisory signals. In: Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, p. 23. ACM (2012)
17. Reuter, T., Papadopoulos, S., Mezaris, V., Cimiano, P.: Reseed: social event detection dataset. In: Proceedings of the 5th ACM Multimedia Systems Conference, pp. 35–40. ACM (2014)
18. Rodriguez, A., Laio, A.: Clustering by fast search and find of density peaks. *Science* **344**(6191), 1492–1496 (2014)
19. Tang, J., Hong, R., Yan, S., Chua, T.S., Qi, G.J., Jain, R.: Image annotation by k nn-sparse graph-based label propagation over noisily tagged web images. *ACM Trans. Intell. Syst. Technol. (TIST)* **2**(2), 14 (2011)
20. Ulges, A., Worring, M., Breuel, T.: Learning visual contexts for image annotation from flickr groups. *IEEE Trans. Multimedia* **13**(2), 330–341 (2011)
21. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3360–3367. IEEE (2010)
22. Wu, Z., Aizawa, K.: Building friend wall for local photo repository by using social attribute annotation. *J. Multimedia* **9**(1), 4–13 (2014)
23. Zhang, D., Islam, M.M., Lu, G.: A review on automatic image annotation techniques. *Pattern Recogn.* **45**(1), 346–362 (2012)