Multimed Tools Appl (2019) 78:13213–13225
https://doi.org/10.1007/s11042-018-5925-5

Image annotation refinement via 2P-KNN based group sparse reconstruction

Qian Ji[1] · Liyan Zhang[2] · Xiangbo Shu[1] · Jinhui Tang[1]

Received: 28 July 2017 / Revised: 12 March 2018 / Accepted: 21 March 2018 /
Published online: 14 April 2018

Abstract Image annotation aims at predicting labels that can accurately describe the semantic information of images. In the past few years, many methods have been proposed to solve the image annotation problem. However, the predicted labels of the images by these methods are usually incomplete, insufficient and noisy, which is unsatisfactory. In this paper, we propose a new method denoted as 2PKNN-GSR (Group Sparse Reconstruction) for image annotation and label refinement. First, we get the predicted labels of the testing images using the traditional method, i.e., a two-step variant of the classical K-nearest neighbor algorithm, called 2PKNN. Then, according to the obtained labels, we divide the K nearest neighbors of an image in the training images into several groups. Finally, we utilize the group sparse reconstruction algorithm to refine the annotated label results which are obtained in the first step. Experimental results on three standard datasets, i.e., Corel 5K, IAPR TC12 and ESP Game, show the superior performance of the proposed method compared with the state-of-the-art methods.

Keywords Image annotation · K nearest neighbor · Group sparsity · Sparse reconstruction

✉ Liyan Zhang
zhangliyan@nuaa.edu.cn

Qian Ji
jqianxixi@163.com

Xiangbo Shu
shuxb@njust.edu.cn

Jinhui Tang
jinhuitang@njust.edu.cn

[1] Nanjing University of Science and Technology, Jiangsu, China

[2] Nanjing University of Aeronautics and Astronautics, Jiangsu, China

1 Introduction

With the prevalence of social networks and digital cameras in our daily life, more and more people share their photos with each other and the number of images presents a trend of the explosive growth on Internet. To efficiently find one image from a mass of images, many tag-based image search engines are emerging rapidly. To enhance the performance of these search engines, more and more researchers study on how to give the accurate labels for images. The aim of image annotation is to assign several labels for an image which can accurately describe the semantic content of the image. Since the existence of the well-known semantic gap [1], image annotation becomes a challenging and difficult task.

Due to the huge amount of images, it is a time-consuming and labor-consuming to manually annotate images. Therefore, in the past decades, many image annotation methods have been proposed to predict the labels for an image, and these methods have achieved better and better annotation results. Because of the simplicity yet effectiveness, the nearest neighbor based image annotation methods have been widely applied into the practice image annotation system. Li et al. proposed an algorithm that accumulated votes from visually similar neighbors to learn tag relevance scalably and reliably [11]. Some methods use the support vector machine (SVM) to annotate images. Tao et al. proposed an asymmetric bagging and random subspace SVM (ABRS-SVM) and combined the random subspace method and SVM to improve the relevance feedback performance [22]. Besides, there are also many other methods for image annotation. Tang et al. proposed a novel generalized deep transfer networks (DTNs), which can transfer label information across heterogeneous domains, textual domain to visual domain and this framework can solve the problem of insufficient training images. [20]. To improve the performance of social image tag refinement, Tang et al. proposed a novel tri-clustered tensor completion framework to collaboratively explore three kinds of information, including users, images and tags [21].

However, the labels obtained by the traditional methods are usually incomplete, insufficient and noisy to accurately describe the semantic content of one image. To solve the problem, many automatic data cleaning methods have been proposed [26]. In this paper, to solve this problem, we propose a novel 2PKNN-GSR method for image annotation and label refinement, which can effectively refine the relevance between the labels and the images to improve the performance of image annotation. The whole framework of the proposed method is illustrated in Fig. 1. The first step of this framework is to obtain the relevance between the labels and the testing images using the traditional 2PKNN method [23]. 2PKNN is a two-step variant of the classical K-nearest neighbor algorithm, and the first step addresses the class-imbalance issue using image-to-label similarity, while the second step use image-to-image to address the issue of weak-labeling. Due to the existence of the above shortcomings, we choose the group sparsity [27] for the group sparse reconstruction to refine the result from the first step. To prove the effectiveness of the proposed method, we conduct the experiments on three well-known datasets, i.e., Corel 5K [3], IAPR TC12 [14] and ESP Game [25] datasets. And experimental results show that the proposed method achieves better performance than the state-of-the-art methods.

Overall, the contributions in this work are summarized as follows:

- We propose a novel 2PKNN-GSR (Group Sparse Reconstruction) method to refine the annotation labels of images, which are obtained by the traditional 2PKNN method.
- In the proposed method, the group sparse reconstruction with a combination of L1 and L2 norms can effectively mine the relevance for image annotation, where L1 norm and

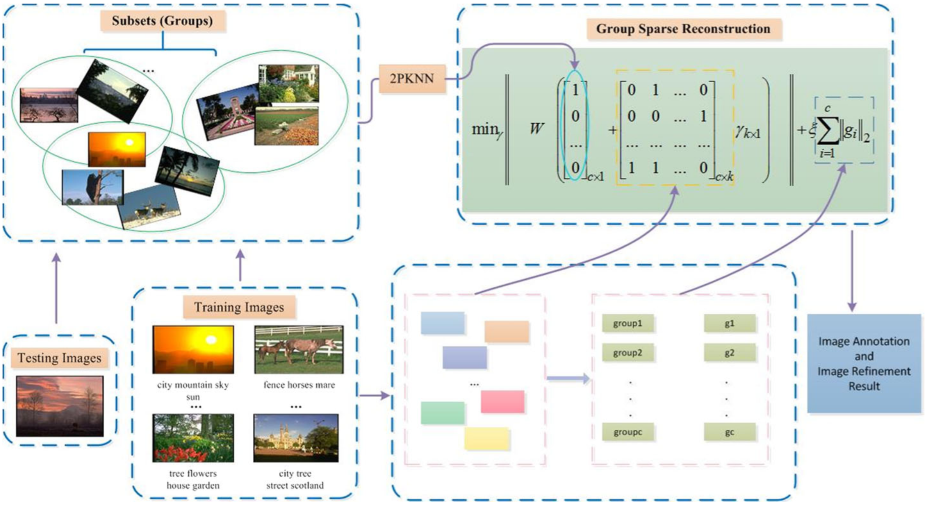

Subsets (Groups)
...
2PKNN
Group Sparse Reconstruction
$\min_{\gamma} \left\| W \begin{bmatrix}1\\0\\\dots\\0\end{bmatrix}_{c\times 1} + \begin{bmatrix}0&1&\dots&0\\0&0&\dots&1\\\dots&\dots&\dots&\dots\\1&1&\dots&0\end{bmatrix}_{c\times k} \gamma_{k\times 1} \right\| + \xi \sum_{i=1}^{c} \|g_i\|_2$
Testing Images
Training Images
city mountain sky sun
fence horses mare
tree flowers house garden
city tree street scotland
...
group1
group2
groupc
g1
g2
gc
Image Annotation and Image Refinement Result

Fig. 1 The framework of our method

L2 norm are used to emphasize the sparsity among the groups and smooth the weights within each group, respectively.

The rest of this paper is organized as follows. Section 2 reviews related works. Section 3 details the proposed method. Experiments are conducted in Section 4, followed by the conclusions in Section 5.

2 Related work

Image annotation has attracted more and more attentions in the field of computer vision and multimedia. Based on the different solutions, existing image annotation methods can be mainly divided into three main categories, including generative models, discriminative models and nearest neighbor based methods. In the past few years, sparsity-based methods have emerged and become more and more popular in image annotation. We will introduce the four categories as follows.

2.1 Generative models

The generative models assign labels to images on the basis of the learnt joint distributions between image features and tags. It can also be interpreted as a collection of mixture models and topic models. As an important part of generative models, mixture models define a joint distribution over image features and labels. Then, the image annotation task is considered as learning the non-parametric density estimators between the images and labels. Lavrenko et al. proposed an approach, called Continuous-space Relevance Model (CRM) [10], whose basic idea is to divide an image into several real-valued feature vectors, compute the joint probability between image features and labels and predict the probability for each label. In the topic models, we regard images as samples from a specific mixture of topics. Blei et al. described a three-level hierarchical Bayesian model, Latent Dirchlet Allocation (LDA) [2].

Putthividhya et al. extended the supervised Latent Dirichlet Allocation to a new probabilistic model called sLDA-bin [17], which can address a multi-variate binary response variable in the annotation data.

2.2 Discriminative models

The discriminative models aim at learning a separate classifier for each class, which can predict the labels class of an image. Szummer et al. previously proposed that images were divided into two categories [18], i.e. image classification. As we all know, the number of an images labels is usually more than two. The problem of the multi-class classification becomes more necessary. To remove the confusing labels, Lavreko et al. presented a model based on an SVM, which modifies the SVM hinge loss function [24]. Recently, Hong et al. proposed a Multiple-Instance Learning (MIL) method which makes use of discriminative feature mapping and feature selection to address the noise of the generated features [8]. In the Cross-Media Relevance Model (CMRM) [9], image annotation is considered as a cross-lingual retrieval problem.

2.3 Nearest neighbor based methods

The nearest neighbor based methods have become more and more popular in the domain of image annotation due to their effectiveness. These methods can be summarized that the labels can be shared among similar images, and various kinds of features are usually combined to compute the distance between images to get the similarities. Makadia et al. proposed a new baseline technique, Joint Equal Contribution (JEC) [14], which considered the image annotation task as a retrieval problem. Guillaumin et al. presented the TagProp [6] method, which learns the weight of each feature group and use label relevance prediction to annotate images. Verma et al. recently proposed 2PKNN [23], which takes advantage of image-to-label and image-to-image similarities to respectively the “class-imbalance” and the “weak-labeling” issues at the same time.

3 The proposed method

In this section, we propose a novel 2PKNN-GSR method for image annotation and label refinement. A summary of the notations in this paper is shown in Table 1.

3.1 Relevance between labels and image

The traditional 2PKNN method can address both the class-imbalance and weak-labeling issues. In this section, we make use of 2PKNN [23] to obtain the relevance between the labels and testing images. The first step of 2PKNN addresses the class-imbalance issue by using image-to-label similarities, while the second step uses image-to-image similarities to address the weak-labeling issue.

Let $X = \{x_1, x_2, ..., x_n\} \in R^{n\times d}$ be a collection of n training images where $x_i \in R^d (i \in [1, n])$ is the ith image. Define $L = \{l_1, l_2, ..., l_c\} \in \{0, 1\}^{n\times c}$ as a dictionary consisting of c labels. The set $T = \{(x_1, t_1), (x_2, t_2), ..., (x_n, t_n)\}$ contains the pairs of the image x_i and its corresponding label set t_i, where $t_i \in \{0, 1\}^c$. And $t_i(j) = 1$ if x_i is annotated by the label l_j and $t_i(j) = 0$ otherwise.

Table 1 Description of symbols

Symbols	Descriptions
X	The collection of training images
x_i	The ith image of the training set
n	The number of the training data
d	The dimension of an image
L	The dictionary of labels
c	The number of the labels
l_i	The ith label in the dictionary
t_i	The label set of image x_i
I	The testing image
T	The set containing the pairs of image x_i and its corresponding label set t_i
T_i	The subset of T containing all the images with the label l_k
$T_{I,i}$	The K1 nearest neighbors of the testing image I in the subsets T_i
T_I	The K1 nearest neighbors of the testing image I in all subsets
$D(I, x_i)$	The visual similarity between I and x_i
P	The relation matrix between the testing images and labels
m	The number of the testing images
$\hat{P}$	The corresponding label matrix of the K2 nearest neighbors of the testing image I in the training set
p	The label vector of a testing image I in P'

According to [23], we can see that the 2PKNN method can solve the problem of weak-labeling and class-imbalance. Define $T_i \subseteq T, \forall i \in \{1, ..., c\}$ as the subset of T that contains all the images with the label $l_k (k \in [1, ...c])$. Given a testing image I, we compute the visual distance between this images and other images in each subset. Due to the more informative diversities of images in each $T_{I,i}$, it is necessary to merge all the subsets to form $T_I = \{T_{I,1} \cup T_{I,2} \cup ... \cup T_{I,c}\} = \cup_{i \in [1,...,c]} T_{I,i}$ as the neighbors of image I. This is the first pass of 2PKNN.

The second pass of 2PKNN is to give the different important values for different labels by assign a weight to each label. Based on the set T_I, given a label l_k, we can write the posterior probability for the testing image I.

$$P(I|l_k) = \sum_{(x_i,t_i) \in T_I} \alpha_{I,x_i} \cdot \beta(l_k \in t_i), \tag{1}$$

where $\alpha_{I,x_i} = \exp(-D(I, x_i))$ is the contribution of the image x_i when we predict a label l_k for I, and $D(I, x_i)$ denotes the visual similarity between I and x_i. And $\beta(l_k \in t_i)$ denotes whether or not the label l_k appears in the set t_i of x_i, namely $\beta(l_k \in t_i) = 1$ if l_k appears in t_i and $\beta(l_k \in t_i) = 0$ otherwise.

Given a testing image I, we can obtain the posterior probability for the label l_k, i.e.,

$$P(l_k|I) = \frac{P(l_k)P(I|l_k)}{P(I)}, \tag{2}$$

According to the 2PKNN method, we can get a relation matrix to shows the relation between the testing images and all the labels. For the following refinement, we choose the M most relative labels for one testing image, and set the entries of the relation matrix according to these labels as 1, while other entries of the relation matrix are set as 0. Then, we define the original relation matrix as $P_{m\times c}$, where m is the number of the testing images.

3.2 Group sparse reconstruction

We can make use of 2PKNN to obtain the relation matrix $P_{m\times c}$, which can achieve image annotation task and show the relation between images and labels. However, these obtained labels are usually insufficient and noisy to describe the whole semantic content of the testing images. To solve the problem, we propose to utilize the group sparse reconstruction [12] to further refine the relation matrix $P_{m\times c}$. The process of group sparse reconstruction can be formulated as follows.

The process of group sparse reconstruction can be formulated as follows

$$\Omega = \min_{\tau} \|W(p - \hat{P}\tau)\|_2^2 + \xi \sum_{i=1}^{c} \|g_i\|_2, \tag{3}$$

where $p \in R^{c\times 1}$ is the label vector of a testing image I in $P'_{m\times c}$, $\hat{P} \in R^{c\times K2}$ is the corresponding label matrix of the K2 nearest neighbors of the image I in the training image set, $\tau \in R^{K2\times 1}$ denotes the weights of each neighbors in the training image set, W is a diagonal matrix defined as $W(i,i) = \exp(p_i)$, and ξ is a tuning parameter to balance the group sparsity. The details of the groups are as follows: First, for each label l_i, all the images in the training images with the label l_i form a group $group_i$. Then, we can define $g_i = \{\tau_{\sigma(i,1)}, \tau_{\sigma(i,2)}, ..., \tau_{\sigma(i,|g_i|)}\}$ as the corresponding index of the K2 nearest neighbors appearing in the g_i in the vector τ.

According to [27], the group sparsity can improve the performance of previous sparse method and ensure more accurate and robust weights. From the latter part in (3), we can see that $\xi \sum_{i=1}^{c} \|g_i\|_2$ is a combination of both L1 and L2 norms. L1 norm is used to emphasize the sparsity among groups, whereas L2 norm is used to smooth the weights within each group.

After getting the optimal parameter ξ, the reconstructed label vector $\hat{p}$ for a testing image I can be obtained as follows.

$$\hat{p} = \hat{P}\tau. \tag{4}$$

Finally, we can use the above-mentioned method for each testing image and further refine the relation matrix P. The algorithm of the proposed 2PKNN-GSR method details in Table 2.

4 Experimental results and analysis

4.1 Datasets

We consider three standard image annotation datasets to evaluate the performance of the proposed 2PKNN-GSR method. Then we compare the performance with previous state-of-the-art methods [4–6, 10, 13–16, 23, 27]. Table 3 shows the information of the three datasets in detail.

Table 2 The algorithm Of the 2PKNN-GSR method

Input: The training image $X \in R^{n \times d}$; the dictionary $L \in R^{n \times c}$; the testing image $I \in R^{1 \times d}$;
Output: Reconstructed label vector $\hat{p}$.

```
1 obtain each subset of training set $T_i \subseteq T, \forall i \in \{1, ..., c\}$;
2 compute the posterior probability for the testing image $I$ given each label
  $l_k (k \in [1, ..., c])$ using (1);
3 assign the importance to each label using (2);
4 for each label $l_k$, all the images in the training image set with the label $l_k$ form a
  group;
5 $g_i = \{\tau_{\sigma(i,1)}, \tau_{\sigma(i,2)}, ..., \tau_{\sigma(i,|g_i|)}\}$;
6 obtain the optimal $\tau$ using (3);
7 obtain the reconstructed label vector using (4).
```

4.2 Features

In the experiments, we use the similar features in [6] and directly merge these features as a set to represent all the images to compare the proposed method with the state-of-the-art methods. Therefore, this set consists of 15 global and local features. The global features contain the Gist and the color histograms in HSV, LAB and RGB. While in the local features, there are the SIFT and hue descriptors, which are obtained from multi-scale grid and Harris-Laplacian interest points. For each image, all the features, except for the Gist, also need to be calculated over three equal horizontal partitions to encode some spatial information of an image. We make use of different measures to compute the distance between different features. For example, L_1 measure for the color histograms in HSV, LAB and RGB, L_2 for the Gist features and χ^2 for SIFT and hue descriptors.

4.3 Evaluation measures

In the experiments, we make use of the annotation precision and recall of each label in the testing image set to evaluate the performance and compare the performance with other state-of-the-art methods. For example, we assume that the number of images annotated by the label l_i in the ground-truth is num_1, and the label l_i is annotated for num_2 images in the testing set where the labels of num_3 images are correct. Therefore, for the label l_i, the precision is defined as $Precision = \frac{num_3}{num_2}$, and the recall is $Recall = \frac{num_3}{num_1}$. Then, according to the precision and recall of each label, we can obtain the average precision P

Table 3 Details for the three datasets used in this work

	Corel 5K	IAPR TC12	ESP Game
No. of images	4999	19627	20770
No. of training images	4500	17665	18689
No. of testing images	499	1962	2081
No. of labels	260	291	268
Labels per image(mean)	3.4	5.7	4.7

and recall R, and further get the percentage $F1-score$ by using $F1 = \frac{2\cdot P\cdot R}{P+R}$. Furthermore, we also compare the value $N+$ which is the number of the labels assigned to at least one testing image. The above parameters can evaluate the performance of the proposed method effectively.

4.4 Results

We firstly evaluate the performance of the proposed method by comparing it with several previous state-of-the-art methods. Table 4 shows the results of the proposed method and other state-of-the-art methods on three datasets (Corel 5K, IAPR TC12 and ESP Game). According to the results, we can conclude that the proposed 2PKNN-GSR method outperforms the previous state-of-the-art methods. From the table, we can see that the recall of the proposed method is the highest on both the IAPR TC12 and ESP Game, the precision is the highest on the Corel 5K dataset, and the $F1-score$ and $N+$ on the three datasets are better than a majority of these methods.

First, we compare the proposed 2PKNN-GSR method with the nearest neighbor based methods [6, 14, 23]. According to the results in Table 4, we can see that the $F1 - score$ of our 2PKNN-GSR is higher than the above three methods. By analyzing the results, we can find that if we only make use of the traditional nearest neighbor based methods, the obtained labels are usually incomplete, inconsistency and error-prone. Since the proposed 2PKNN-GSR method uses the group sparse reconstruction to refine the annotation results which are obtained by traditional image annotation methods, it significantly outperforms the nearest neighbor based methods.

Then, we also compare our method with Sparsity-based methods [15, 27].According to the obtained results in Table 4, we can see that the proposed 2PKNN-GSR method achieves the better performance than the above two sparsity-based methods. The disadvantages of these sparsity-based methods are that they use all data to train the model and do not ignore the useless data. Therefore, it may cause the data redundancy. It is noticed that the proposed 2PKNN-GSR method firstly uses 2PKNN method to obtain the neighbors of the testing image, which can avoid the data redundancy as much as possible to improve the image annotation results.

Moreover, to intuitively show the effectiveness of the proposed 2PKNN-GSR method, we also some example results of the proposed method on the Corel 5K dataset, as shown in Fig. 2. For example, the predicted labels of the fifth image include the word "tree" yet the ground truth do not, and the word can describe the image better. The ground truth of the last image includes the word "house", yet it cannot describe the image. However, the proposed method refines the results. The predicted labels delete "house" and add two words, "sky" and "water", which can show this image completely.

We can analyze the proposed 2PKNN-GSR method from following two aspects:

The first step of our method is to get the predicted labels of the testing images using the traditional 2PKNN method. This 2PKNN method consists of two steps, which respectively use "image-to-label" similarity and "image-to-image" similarity, and can effectively address the issues of weak-labeling and class-imbalance.

However, the predicted labels obtained from the traditional methods usually are incomplete, insufficient and noisy to describe the semantic content accurately. To solve the problem, in the second step, we propose the 2PKNN-GSR method to take advantage of the group sparse reconstruction to refine the results obtained by the first step. In the group structure, both L1 and L2 norms are combined, and L1 norm is used to emphasize the sparsity among the groups, while L2 norm is to smooth the weights within each group.

Table 4 Details for three datasets (corel 5K, IAPR TC12 and ESP game)

	Corel 5K				IAPR TC12				ESP Game			
Method	P	R	F	N+	P	R	F	N+	P	R	F	N+
CRM [10]	16	19	17.4	107	-	-	-	-	-	-	-	-
MBRM [4]	24	25	24.5	122	24	23	23.5	223	18	19	18.5	209
TGLM [13]	25	29	26.9	131	-	-	-	-	-	-	-	-
JEC [14]	27	32	29.3	139	28	29	28.5	250	22	25	23.4	224
GS [27]	30	33	31.4	146	32	29	30.4	252	-	-	-	-
RF-opt [5]	29	40	33.6	157	44	31	36.4	253	41	26	31.8	235
CCD [16]	36	41	38.3	159	44	29	35.0	251	36	24	28.8	232
TagProp [6]	33	42	37.0	160	46	35	39.8	266	39	27	31.9	239
SKL-CRM [15]	39	46	42	184	47	32	38	274	41	26	32	248
2PKNN [23]	39	40	39.5	177	49	32	38.7	274	51	23	31.7	245
2PKNN-GSR	39.5	43.5	41.4	189	42.6	39.5	41.0	279	41	29.5	34.3	251

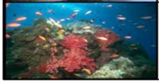
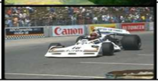
Image	Ground truth	Predicted labels	Image	Ground truth	Predicted labels
	sky sun clouds	sky sun water clouds sunset		sky jet plane smoke	sky clouds jet plane smoke
	wall cars tracks formula	water wall cars tracks formula		tree buildings street sculpture	sky water sterr grass buildings
	field horse mare foals	tree field horses mare foals		water bear black river	water bear snow black birds
	people hut	water tree people grass buildings		water rocks eat tiger	water tree people cat tiger
	sky sun clouds	sky sun water clouds sunset		tree flowers house garden	sky water tree flowers garden
	coral fish ocean reefs	water coral fish ocean reefs		wall cars tracks formula	wall cars tracks people formula

Fig. 2 Details For the Three Datasets Used in This Work

All of the above, we can summarize the novelty and technical contribution of the proposed 2PKNN-GSR method. The method combined the traditional 2PKNN method with the group sparse reconstruction effectively. First, we use the traditional 2PKNN to obtain the relation between the testing images and labels. However, these labels are usually incomplete, insufficient and noisy. Then, the proposed 2PKNN-GSR method refined the results by the group sparse reconstruction, which uses both L1 and L2 norms at the same time. Finally, we can improve the image annotation results.

5 Conclusion and future work

In this paper, we propose a novel method, 2PKNN-GSR, to refine image annotation results by using the group sparse reconstruction to improve the performance. First, we get the predicted labels of the images using the traditional 2PKNN method, which make use of “image-to-label” and “image-to-image” similarities to address the weak-labeling and class-imbalance issues. However, since the predicted labels of the images are usually incomplete, insufficient and noisy to describe the whole semantic content of images accurately, thus causing the unsatisfactory results. Then, we take advantage of the group sparse reconstruction to refine the above results obtained by 2PKNN. We conduct the experiments and theoretical analysis on three standard datasets, i.e. Corel 5K dataset, IAPR TC12 dataset and ESP Game dataset. Experimental results on the three datasets show that the proposed 2PKNN-GSR method outperforms the several previous state-of-the-art methods in the annotation quality. In the future, we plan to select “clean” samples for learning recognizer to prove the effectiveness of the proposed 2PKNN-GSR method and improve the image annotation performance by adopting more efficient refinement procedure.

References

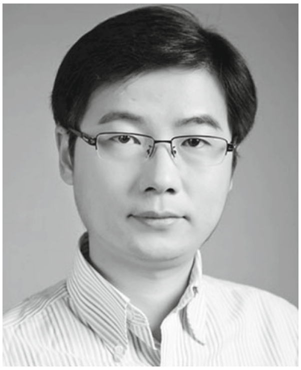
1. Bahmanyar R, Ambar MMD, Datcu M (2015) The semantic gap: an exploration of user and computer perspectives in earth observation images. IEEE Geosci Remote Sens Lett 12(10):2046–2050
2. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. Journal of Machine Learning Research 3:993–1022
3. Duygulu P, Barnard K, de Freitas JF, Forsyth DA (2002) Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. European Conference on Computer Vision 4:97–112
4. Feng S, Manmatha R, Lavrenko V (2004) Multiple Bernoulli relevance models for image and video annotation. Comput Vis Pattern Recognit 2:1002–1009
5. Fu H, Zhang Q, Qiu G (2012) Random forest for image annotation. European Conference on Computer Vision 2:86–99
6. Guillaumin M, Mensink T, Verbeek J, Schmid C (2009) Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In: IEEE 12th International Conference on Computer Vision. IEEE, pp 309–316
7. Han Y, Wu F, Tian Q, Zhuang Y (2012) Image Annotation by InputCOutput Structural Grouping Sparsity. IEEE Trans Image Process 21(6):3066–3079
8. Hong R, Wang M, Gao Y, Tao D, Li X, Wu X (2014) Image annotation by multiple-instance learning with discriminative feature mapping and selection. IEEE Trans Cybern 44(5):669–680
9. Jeon J, Lavrenko V, Manmatha R (2003) Automatic image annotation and retrieval using cross-media relevance models. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, pp 119–126
10. Lavrenko V, Manmatha R, Jeon J (2004) A model for learning the semantics of pictures. In: Advances in Neural Information Processing Systems, vol 16, pp 553–560
11. Li X, Snoek CGM, Worring M (2008) Learning tag relevance by neighbor voting for social image retrieval. Proceedings of 1st ACM international conference on multimedia information retrieval. ACM, pp 180–187
12. Lin Z, Ding G, Hu M, Wang J, Ye X (2013) Image tag completion via image-specific and tag-specific linear sparse reconstructions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1618–1625
13. Liu J, Li M, Liu Q, Lu H, Ma S (2009) Image annotation via graph learning. Pattern Recogn 42(2):218–228
14. Makadia A, Pavlovic V, Kumar S (2008) A new baseline for image annotation. European Conference on Computer Vision 3:316–329
15. Moran S, Lavrenko V (2014) Sparse kernel learning for image annotation. Proceedings of international conference on multimedia retrieval, pp 113–120
16. Nakayama H (2011) Linear distance metric learning for large-scale generic image recognition. PhD thesis, The University of Tokyo
17. Putthividhya D, Attias HT, Nagarajan SS (2010) Supervised topic model for automatic image annotation. IEEE International Conference on Acoustics, Speech, & Signal Processing 1:1894–1897
18. Szummer M, Picard R (1998) Indoor-outdoor image classification. In: Proceedings of IEEE international workshop on Contentbased Access of Image and Video Database, pp 42–51
19. Tang J, Hong R, Yan S, Chua TS, Qi GJ, Jain R (2011) Image annotation by knn-sparse graph-based label propagation over noisily tagged web images. ACM Trans Intell Syst Technol 2(2):1–15
20. Tang J, Shu X, Qi G, Li Z, Wang M, Yan S, Jain R (2016) Generalized Deep Transfer Networks for Knowledge Propagation in Heterogeneous Domains. CM Trans Multimed Comput Commun Appl 12(4s):68
21. Tang J, Shu X, Qi G, Li Z, Wang M, Yan S, Jain R (2016) ri-Clustered Tensor Completion for Social-Aware Image Tag Refinement. IEEE Transactions on Pattern Analysis Machine Intelligence. pp(99), pp 1-1
22. Tao D, Tang X, Li X, Wu X (2006) Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Trans Pattern Anal Mach Intell 28(7):1088–1099
23. Verma Y, Jawahar C (2012) Image annotation using metric learning in semantic neighborhoods. European Conference on Computer Vision 3:836–849
24. Verma Y, Jawahar C (2013) Exploring SVM for image annotation in presence of confusing labels. British Machine Vision Conference 1:1–11
25. Von Ahn L, Dabbish L (2004) Labeling images with a computer game. In: SIGCHI Conference on Human Factors in Computing Systems, pp 319–326

26. Yu J, Rui Y, Tao D (2014) Click Prediction for Web Image Reranking Using Multimodal Sparse Coding. IEEE Trans Image Process 23(5):2019–2032
27. Zhang S, Huang J, Huang Y, Yu Y, Li H, Metaxas DN (2010) Automatic image annotation using group sparsity. Comput Vis Pattern Recognit 3:3312–3319

Qian Ji received the BS degree in School of Computer Science and Engineering from Nanjing University of Science and Technology in 2015. She is currently a candidate for a Ph.D. degree in School of Computer Science and Engineering, Nanjing University of Science and Technology, China. Her research interests include computer vision and information retrieval.

Liyan Zhang received the PhD degree in computer science from the University of California, Irvine, in 2014. She is currently an associate professor in the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics. Her research interests include multimedia analysis and computer vision. She has received the Best Paper Award in ICMR 2013 and the Best Student Paper Awards in MMM 2016 and ICIMCS 2017.

Xiangbo Shu received the PhD degree from Nanjing University of Science and Technology, in July 2016. He is an assistant professor in the School of Computer Science and Engineering, Nanjing University of Science and Technology, China. From 2014 to 2015, he worked as a visiting scholar in the Department of Electrical and Computer Engineering, National University of Singapore. His research interests include computer vision and machine learning. He has received the Best Student Paper Award in MMM 2016 and the Best Paper Runner-up in ACMMM 2015.



Jinhui Tang received the BE and PhD degrees both from the University of Science and Technology of China, in July 2003 and July 2008, respectively. He is a professor in the School of Computer Science and Engineering, Nanjing University of Science and Technology, China. From 2008 to 2010, he worked as a research fellow in the School of Computing, National University of Singapore. His current research interests include large-scale multimedia search. He has authored more than 100 journal and conference papers in these areas. He received the ACM China Rising Star Award and a co-recipient of the Best Student Paper Award in MMM 2016, and Best Paper Award in ACM MM 2007, PCM 2011, and ICIMCS 2011. He is a senior member of the IEEE.