International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 8, No. 4 (2015), pp. 145-154
http://dx.doi.org/10.14257/ijsip.2015.8.4.13

Cross-Media Retrieval using Probabilistic Model of Automatic Image Annotation

Ying Xia, YunLong Wu and JiangFan Feng

Research Center of Spatial Information System, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
xiaying@cqupt.edu.cn, wylcqupt@outlook.com

Abstract

In recent years, automatic image annotation (AIA) has been applied to cross-media retrieval usually due to its advantage of mining correlations of images and annotation texts efficiently. However, some AIA methods just annotate images as a unit and the accuracy of annotation may not be acceptable. In this paper, we propose a kind of probabilistic model which may assign keywords to an un-annotated image automatically based on a training dataset of images. Images in the training dataset are segmented into regions and a kind of vocabulary called blob is used to represent these image regions. Blobs are generated by using K-Means algorithm to cluster these image regions. Through this model, we can predict the probability of assigning a keyword into a blob. After the accomplishment of annotation, a keyword corresponds to one image region. Furthermore, the feature vectors of text documents are generated by TF.IDF method and images' automatic annotation information is used to retrieve relevant text documents. Experiments on the IAPR TC-12 dataset and 500 Wikipedia webpages about landscape show the usefulness of applying probabilistic model of AIA to the cross-media retrieval.

Keywords: *automatic image annotation, cross-media retrieval, probabilistic model*

1. Introduction

With the tremendous growth of multimedia data, content-based retrieval is proposed to search this multi-modality information precisely. Zakariya S M and Chen refer to content-based image retrieval (CBIR) using clustering in [1, 2]. Simon Tong and Edward Chang [3] concern active learning algorithm, exactly support vector machine (SVM), to conduct effective image retrieval. Peker, K. A [4] extracts 128-D binary vectors of SIFT features to compute distance between images. E. Wold, T. Blum, D. Keislar and J. Wheaten [5] take advantage of the loudness, pitch, brightness and bandwidth of sound to measure the audio data's correlations. S. Ghodeswar and B. B. Meshram [6] talk about video segmentation, key frame selection and feature extraction. In addition, to bridge the semantic gap more efficiently, relevance feedback (RF) is considered as a significant way to discover multimedia data's semantics in [7-9]. However, most methods of content-based retrieval only focus on features of single modality data. Correlations among multimedia data with different modalities are ignored, so cross-media retrieval is put forward to solve this problem. Zhuang and Wu [10, 11] introduce an isomorphic subspace constructed based on Canonical Correlation Analysis (CCA) to learn cross-modal correlations of multimedia data. B. Lu, G. R. Wang and Y. Yuan [12] propose a multi-modality semantic relationship graph (MSRG) to map media objects onto an isomorphic semantic space and an efficient indexing MK-tree to manage the media objects. These methods solve the problems of cross-media to some extents, but a great deal of annotation information may be ignored. In current time, there exists myriad variety of Web images with manual annotations, such as anchor text and labels. Texts are very significant for mining the correlations between images and texts. As a result, to avoid the labor intensive procedure and improve the efficiency, automatic image annotation becomes an appropriate way to mine the correlations between images and texts.

The most intrinsic problem for automatic image annotation is how to improve the accuracy of annotation. S. L. Feng, R. Manmatha and V. Lavrenko [13] use a multiple Bernoulli relevance model to do both automatic image annotation and one word queries retrieval. A. Makadia, V. Pavlovic and S. Kumar [14] introduce a new baseline technique for image annotation that treats annotation as a retrieval problem. They utilize low-level image features and a simple combination of basic distances to find nearest neighbors of a given image. M. Guillaumin, T. Mensink, J. Verbeek and C. Schmid [15] use a weighted nearest-neighbor model to exploit labeled training images, thus they combining a collection of image similarity metrics that cover different aspects of image content. Moreover, they also introduce a word specific sigmoidal modulation of the weighted neighbor tag predictions to boost the recall of rare words. S. Zhang, J. Huang, Y. Huang and Y. Yang introduce a regularization based feature selection algorithm to leverage both the sparsity and clustering properties of features, and incorporate it into the image annotation task and they also propose a novel approach to iteratively obtain similar and dissimilar pairs from both the keyword similarity and the relevance feedback in [16]. D. D. Burdescu, C. G. Mihai, L. Stanescu and M. Brezovan [17] present a system used in the medical domain for three tasks: image annotation, semantic based image retrieval and content based image retrieval. J. Y. Pan, H. J. Yang, C. Faloutsos and P. Duygulu [18] propose a novel, graph-based approach to discover multi-modal correlations. Researchers take advantage of automatic image annotation to mine the correlations among media objects of different modalities also in [19, 20]. However, in these AIA methods keywords are just assigned to the entire image as a unit and higher semantic information are lost.

This paper proposes a probabilistic model of automatic image annotation to research the correlation between images and texts. In the model, keywords can be assigned to the un-annotated image, also there exists one correspondence between a keyword and an image region. After these procedures accomplished, images can be annotated in acceptable precision. Moreover, assigning keyword to every image region contributes to taking advantage of other high-level semantic information for cross-media retrieval, like spatial relations among image regions.

2. Features of Images

Before assigning keywords to the image regions, every image in training dataset should be segmented into several regions using image’s low-level features. Then to generate a set of numbers corresponding to regions to represent every image, these regions should be clustered into a unique number of region categories. Similarly, for every query image in test dataset, it can be segmented into image regions. These regions can be classified using cluster space of training dataset.

Duygulu and Barnard [21] propose a model of object recognition used to annotate image regions with keywords. They apply Normalized Cuts [22] to segment images into discrete regions and afterwards they cluster these image regions into special image vocabularies called blobs using K-Means algorithm. So, a blob vocabulary corresponds to one kind of image regions and is given a unique number as its identification number. Duygulu and Barnard [21] propose to utilize EM algorithm to predict probability of adding keywords to a new image’s blobs. However, image segmentation is a very erroneous process and results are not satisfactory in general. Although algorithms of image segmentation need to be improved, the generation of blobs is still significant for automatic image annotation. In this paper, we are not focused on image segmentation algorithm, but the probabilistic model and its result of automatic annotation and cross-media retrieval. We use the Segmented and Annotated IAPR-12 dataset in which images are segmented and annotated manually.

3. A Probabilistic Model

In this section, we illustrate the probabilistic model of AIA first. Actually, it is the improvement of cross-media relevance model (CMRM) [19]. Different from CMRM assigning keywords into the whole images, this probabilistic model can associate annotated keyword with every image region. We assume that there exists a training dataset T , in which every image $K \in T$ has been segmented and annotated, that is, K has a set of blobs $\{b_1 \cdots b_m\}$ and a set of words $\{w_1 \cdots w_n\}$: $K = \{b_1 \cdots b_m, w_1 \cdots w_n\}$. In order to simplify the statistics of blobs, we need to make a necessary assumption. Instead of assuming that there is a correspondence between a blob $b_m \in K$ and a word $w_n \in K$, we assume that a set of blobs $\{b_1 \cdots b_m\}$ is related to a set of words $\{w_1 \cdots w_n\}$. Given an un-annotated test image I , which has only a set of blobs $\{b_1 \cdots b_z\}$, the probabilistic model of automatic annotation is formulated as follows.

[23] Introduces a variety of information retrieval models, especially probabilistic model, thus we assume that there exists some probability distribution $P(\bullet \mid I)$ for image I . [24] leverages the probability distribution $P(\bullet \mid I)$ as the relevance model of I . Besides, for the blob b in the image I ($b \in I$), [21] introduces the probability distribution $P(\bullet \mid b)$. Although the model in [21] may not access to good results, it provides a useful method. The relevance model can be thought as a sample space in which the sample point is one of all possible keywords which could be assigned to the blob b .

Based on this relevance model $P(\bullet \mid b)$, the process of annotating the image $I = \{b_1 \cdots b_z\}$ is actually sampling z words from the model for every blob b . So the probability of any word appearing in the training dataset T can be calculated when sampled from $P(\bullet \mid b)$ and then we need to estimate the probability $P(w \mid b)$ for every word w over all blobs in image I . For any blob b in image I , the probability of assigning a keyword into a blob is approximated by the conditional probability of w given b :

$$P(w \mid b) = \frac{P(w,b)}{P(b)} \tag{1}$$

We compute the joint probability of the word w and the blob b in the same image using the training dataset T . The joint distribution can be calculated as the expectation over images K in the training dataset T :

$$P(w,b) = \sum_{K \in T} P(K) P(w,b \mid K) \tag{2}$$

We assume that variables w and b are independent, so are the conditional probabilities $P(w \mid K)$ and $P(b \mid K)$. Equation (2) can be represented as follows:

$$P(w,b) = \sum_{K \in T} P(K) P(w \mid K) P(b \mid K) \tag{3}$$

The prior probabilities $P(K)$ are kept uniform over all images in T . The probability of drawing the word w or the blob b from relevance model of K is formulated as:

$$P(w \mid K) = (1 - \alpha_K) \frac{N(w,K)}{|K|} + \alpha_K \frac{N(w,T)}{|T|} \tag{4}$$

$$P(b \mid K) = (1 - \beta_K) \frac{N(b,K)}{|K|} + \beta_K \frac{N(b,T)}{|T|} \tag{5}$$

In formula (4) and (5), $N(w,K)$ refers to the actual number of times that the word w is attached to the image K. $N(w,T)$ describes the total number of times that the word w occurs in the training dataset T. Similarly, $N(b,K)$ denotes the actual number of times that the blob appears in the image K and $N(b,T)$ is the number of occurrences of blob b in dataset T. $|K|$ represents count of occurrences of words and blobs occurring in the image $K \in T$ and $|T|$ stands for total size of training dataset T .The smoothing parameters α_K and β_K determine the degree of interpolation between the maximum likelihood estimates and the background probabilities for words and the blobs.

According to formula (1)-(5), over all words appearing in the training dataset T for every blob $b \in \{b_1 \square\square b_z\}$ in image I , $P(w \mid b)$ can be calculated and ranked. Then we can assign the keyword having highest probability to the corresponding blob b . Thus we predict a desired number n of keywords to annotate the image I . After annotation accomplished, annotation information can be applied to retrieval task.

In cross-media retrieval, we can use one modal data to retrieve another modal data. In this paper, firstly, we assign keywords to an un-annotated image automatically, extract the features of text in the webpages by TF-IDF method, and then utilize the annotation information of images to retrieve text. We are not concentrated on details of retrieval, so we do not set the special weight of query keywords in the text retrieval but only according to values generated by TF-IDF.

4. Experimental Results

In this section the experimental datasets are discussed first, and then experimental results are showed through comparing different methods. Finally the application results of text retrieval using image query are presented.

4.1. Datasets

In this paper, the accuracy of probabilistic model is emphasized, and it can be illustrated by the precision and recall, so two kinds of dataset are selected in the experiments. In the first dataset called Segmented and Annotated IAPR TC-12, each image has been manually segmented and the resultant regions have been annotated according to a hierarchy of concepts, like entity-landscape-nature-vegetation. Visual features such as region area, width and height have also been extracted from each region. We select 1500 images as training dataset and 300 images as test dataset in the IAPR TC-12 for automatic image annotation experiment. For images in the test dataset, the original annotation information will be ignored when annotating images. After accomplishment of annotation, the original annotation information is used to test the accuracy of image annotation. The second dataset, consisted of text documents selected from 500 Wikipedia webpages about landscape, can be used for a kind of cross-media retrieval, that is, text retrieval using image query.

4.2. Automatic Image Annotation Results and Analysis

The probabilistic model can be applied to annotate images. To evaluate the accuracy of this model, single keyword appearing in the training dataset will be used to retrieve images in the test dataset (note that this is not ranked retrieval). We can evaluate the performance of the model by the precision and recall. The precision denotes the number of correctly retrieved images divided by the number of retrieved images, and the recall represents the number of correctly retrieved images divided by the number of relevant images in the test dataset. Probabilistic model has two smoothing parameters, α_K and β_K .

These parameters are estimated through the training dataset. We concentrate on the average precision and recall to evaluate different models. To pick the best parameters out, we use the F-measure.

$$F = \frac{2 * recall * precision}{recall + precision}$$

As Table 1 and Table 2 show, the values of parameters range from 0.1 to 0.9 for judging the experiment's result. Every group of values is iterated using F-measure, and results show that the group of $\alpha_K = 0.1$ and $\beta_K = 0.5$ is the best choice.

Table 1. Average Precision of Different Parameters

Average Precision	α_K =0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
β_K =0.1	0.12	0.08	0.10	0.15	0.07	0.09	0.10	0.25	0.14
β_K =0.2	0.19	0.11	0.08	0.23	0.10	0.09	0.29	0.13	0.06
β_K =0.3	0.22	0.12	0.12	0.23	0.12	0.15	0.25	0.14	0.08
β_K =0.4	0.31	0.08	0.32	0.23	0.28	0.23	0.19	0.06	0.12
β_K =0.5	0.35	0.10	0.15	0.16	0.15	0.25	0.15	0.08	0.07
β_K =0.6	0.30	0.19	0.16	0.18	0.15	0.32	0.10	0.08	0.26
β_K =0.7	0.14	0.18	0.08	0.18	0.15	0.18	0.09	0.14	0.22
β_K =0.8	0.11	0.20	0.12	0.09	0.12	0.10	0.10	0.17	0.22
β_K =0.9	0.13	0.15	0.10	0.10	0.09	0.10	0.17	0.18	0.12

Table 2. Average Recall of Different Parameters

Average Recall	α_K =0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
β_K =0.1	0.12	0.08	0.10	0.15	0.12	0.09	0.12	0.30	0.14
β_K =0.2	0.21	0.15	0.08	0.23	0.10	0.09	0.29	0.13	0.10
β_K =0.3	0.22	0.12	0.20	0.30	0.15	0.20	0.30	0.14	0.12
β_K =0.4	0.34	0.10	0.32	0.23	0.28	0.23	0.20	0.12	0.16
β_K =0.5	0.44	0.10	0.15	0.16	0.15	0.30	0.15	0.08	0.07
β_K =0.6	0.35	0.19	0.16	0.20	0.19	0.32	0.10	0.12	0.26
β_K =0.7	0.18	0.22	0.10	0.22	0.20	0.18	0.14	0.20	0.24
β_K =0.8	0.11	0.25	0.12	0.18	0.14	0.10	0.10	0.21	0.22
β_K =0.9	0.19	0.15	0.10	0.10	0.09	0.15	0.24	0.18	0.20

Here we compare the results of six models, the cross-media relevance model (CMRM) [19], the multiple-bernoulli relevance model (MBRM) [13], the least absolute shrinkage and selection operator (LASSO) [14], the tag relevance prediction model (TagProp) [15], the group sparsity (GS) [16] and probabilistic model. All the models annotate images with different numbers of keywords according to the number of image's blobs. There are total 73 kinds of annotation words in the test dataset and 20 high-frequency keywords are selected for image retrieval. Figure 1 and Figure 2 show the precision and recall using a set of high-frequency keywords as single word query. MBRM, LASSO, TagProp and GS concentrate on low-level features, so these models can get good results with appropriate low-level features. Thus, the effect of these models depends on the extraction of low-level features. CMRM and probabilistic model are probability-based models. The size of training dataset and the chosen distribution make a difference to the performance of models.

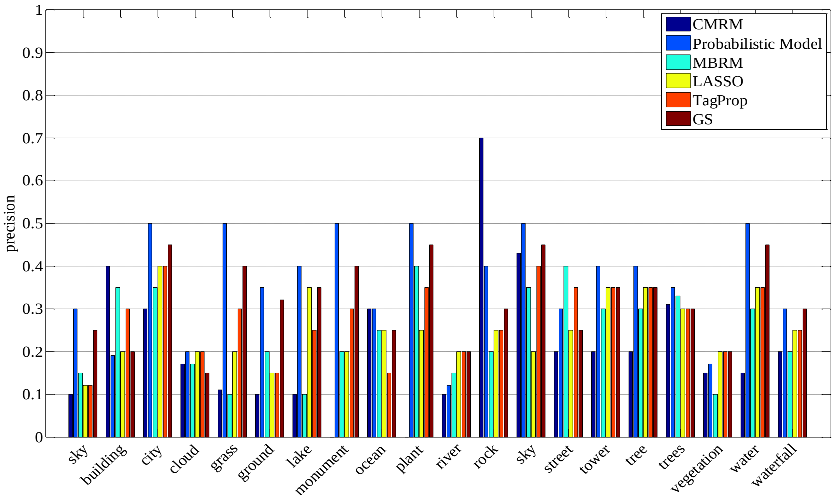


Figure 1. Precision of Image Retrieval using Single Keyword

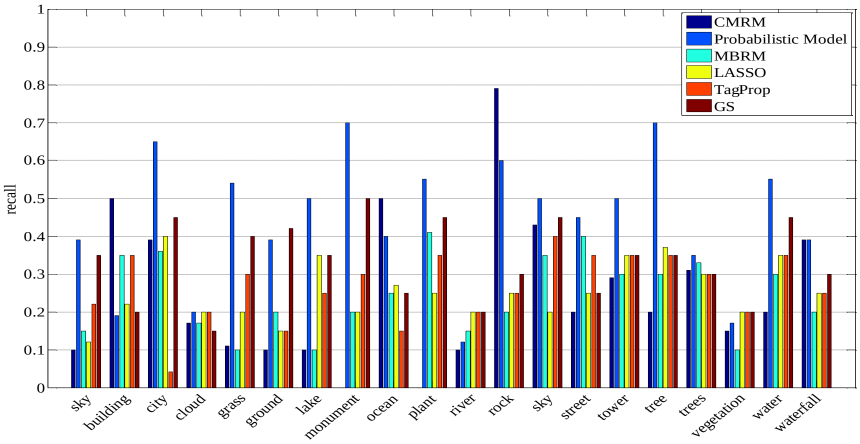


Figure 2. Recall of Image Retrieval using Single Keyword

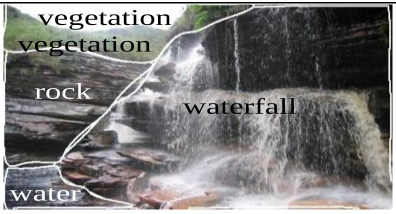
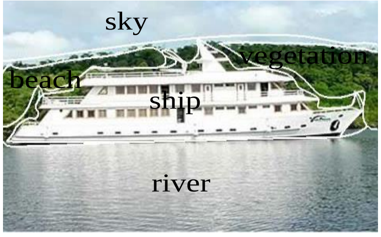
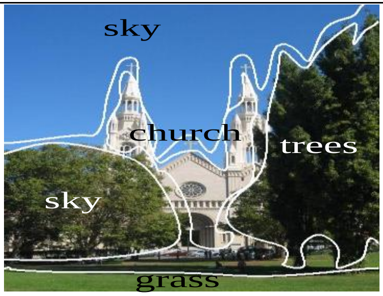
Table 3 shows the average precision and recall of these six models. Obviously, in terms of recall and precision, our approach is best to some extents and GS model also get reasonable results than other four models. Probabilistic model estimate the joint probability $P(w,b)$, so it outperform CMRM. GS utilize both the sparsity and clustering to get better low-level features, so it is better than MBRM, LASSO and TagProp.

Table 3. Average Precision and Recall

Annotation Method	Average Precision	Average Recall
CMRM	0.22	0.32
Probabilistic Model	0.35	0.44
MBRM	0.24	0.25
LASSO	0.25	0.28
TagProp	0.27	0.35
GS	0.32	0.33

In addition, three examples of automatic image annotation are showed using probabilistic model. Table 4 shows that every keyword can be annotated into a blob. By comparing true annotation with automatic annotation, we can get the conclusion that probabilistic model of automatic image annotation can mine relatively correct semantic information, but it depends on the performance of image segmentation.

Table 4. Examples of Image Annotation

image	True Annotation	Automatic Annotation
	sky, water, waterfall, rock, vegetation	
	river, ship, vegetation, sky	
	grass, trees, sky, church	

4.3. Ranked Text Retrieval

Here probabilistic model are evaluated through a kind of cross-media retrieval, that is, text retrieval using image query. The dataset of images used to retrieve text is the test dataset annotated automatically using probabilistic model, and the value of parameters α_K is 0.1 and β_K is 0.5. The dataset of retrieved text consists of 500 webpages and we are only focus on text information. For the text information, the TF-IDF method is used to extract feature vector.

$$TF = \frac{term\_num}{total\_num}$$

$$IDF = \log \frac{corpus\_size}{doc\_num + 1}$$

$$TF - IDF = TF \times IDF$$

Here, $term\_num$ denotes the number of unique word w appearing in one document. $total\_num$ represents the total number of words in one document. $corpus\_size$ reflects the number of documents in the corpus and doc_num is the number of documents containing the unique word w. In the process of retrieval task, we concentrate on these automatically annotated keywords of all images in test dataset and generate the TF-IDF values to retrieve relevant text. The TF-IDF values can be calculated and ranked, so the text retrieval task is a ranked process.

We also compare the results of six models according to the recall and precision, but we have to redefine them. Here, the precision is the number of correctly retrieved text documents divided by the number of retrieved text documents, and the recall is the number of correctly retrieved text documents divided by the number of relevant text documents in the corpus. Table 5 shows the results of six models in the text retrieval. It demonstrates that average precision and recall depend on the accuracy of automatic image annotation. Probabilistic model just annotate images with keywords. To retrieve text more precisely, higher level feature information must be extracted.

Table 5. Results of Text Retrieval using Six Models

Annotation Method	Average Precision In Text Retrieval	Average Recall In Text Retrieval
CMRM	0.29	0.35
Probabilistic Model	0.37	0.44
MBRM	0.22	0.27
LASSO	0.26	0.31
TagProp	0.30	0.36
GS	0.31	0.32

5. Conclusion and Future Work

In this paper, we proposed a novel approach for automatic image annotation. Then the results of automatic annotation information are used to retrieve relevant text. Through comparing the results of experiments we have shown that the probabilistic model has better performance for automatic image annotation and cross-media retrieval. It can gain higher precision and recall for text retrieval using image query. On the other hand, although probabilistic model has promising results, the parameters of probabilistic model must be hand-tuned. Moreover, the performance of the model when changing these parameters is needed to detect. The future work can include the automation of parameter selection.

Acknowledgements

This work was financially supported by the Natural Science Foundation of China (41201378), Natural Science Foundation Project of Chongqing CSTC (cstc2012jjA40014), and Doctor-al Startup Foundation of Chongqing University of Posts and Telecommunications (A2012-34).

References

[1] S. M. Zakariya, R. Ali and N. Ahmad, “Combining visual features of an image at different precision value of unsupervised content based image retrieval”, Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on IEEE, **(2010)**, pp. 1-4.

[2] Y. Chen, J. Z. Wang and R. Krovetz, “CLUE: Cluster-Based Retrieval of Images by Unsupervised learing. IEEE Transaction on Image Processing, vol. 14, **(2005)** August, pp. 1187-1201.

[3] S. Tong and E. Chang, “Support vector machine active learning for image retrieval”, Proceedings of the ninth ACM international conference on Multimedia. ACM, **(2001)**, pp. 107-118.

[4] K. A. Peker, "Binary SIFT: Fast image retrieval using binary quantized SIFT features. Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on IEEE, **(2011)**, pp. 217-222.

[5] E. Wold, T. Blum, D. Keislar and J. Wheaten, "Content-based classification, search and retrieval of audio", MultiMedia, IEEE, **(1996)**, pp. 27-36.

[6] S. Ghodeswar and B. B. Meshram, "Content Based Video Retrieval", Proceedings of ISCET, **(2010)**, pp. 135.

[7] Y. Rui, T. S. Huang, M. Ortega and S. Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval", Circuits and Systems for Video Technology, IEEE Transactions on, **(1998)**, pp. 644-655.

[8] X. He, W. Y. Ma and H. J. Zhang, "Learning an image manifold for retrieval", Proceedings of the 12th annual ACM international conference on Multimedia. ACM, **(2004)**, pp. 17-23.

[9] J. He, M. Li, H. J. Zhang, H. Tong and C. Zhang, "Manifold-ranking based image retrieval", Proceedings of the 12th annual ACM international conference on Multimedia. ACM, **(2004)**, pp. 9-16.

[10] F. Wu, H. Zhang and Y. Zhuang, "Learning semantic correlations for cross-media retrieval", Image Processing, 2006 IEEE International Conference on IEEE, **(2006)**, pp. 1465-1468.

[11] H. Zhang, Y. Zhuang and F. Wu, "Cross-modal correlation learning for clustering on image-audio dataset", Proceedings of the 15th international conference on Multimedia. ACM, **(2007)**, pp. 273-276.

[12] B. Lu, G. R. Wang and Y. Yuan, "A novel approach towards large scale cross-media retrieval", Journal of Computer Science and Technology, vol. 27, no. 6, **(2012)**, pp. 1140-1149.

[13] S. L. Feng, R. Manmatha and V. Lavrenko, "Multiple Bernoulli relevance models for image and video annotation", In CVPR, **(2004)**.

[14] A. Makadia, V. Pavlovic and S. Kumar, "A new baseline for image annotation", In ECCV, **(2008)**, pp. 316-329.

[15] M. Guillaumin, T. Mensink, J. Verbeek and C. Schmid, "Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation", In ICCV, **(2009)**.

[16] S. Zhang, J. Huang, Y. Huang and Y. Yang, "Automatic image annotation using group sparsity", In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on IEEE, **(2010)**, pp. 3312-3319.

[17] D. D. Burdescu, C. G. Mihai, L. Stanescu and M. Brezovan, "Automatic image annotation and semantic based image retrieval for medical domain", Neurocomputing, **(2013)**, pp. 33-48.

[18] J. Y. Pan, H. J. Yang, C. Faloutsos and P. Duygulu, "Automatic multimedia cross-modal correlation discovery", Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, **(2004)**, pp. 653-658.

[19] J. Jeon, V. Lavrenko and R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models", Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. ACM, **(2003)**, pp. 119-126.

[20] J. Guo and X. Liao, "Cross-Media Image Retrieval via Latent Semantic Indexing and Mixed Bagging", Computer Science and Information Engineering, 2009 WRI World Congress on IEEE, **(2009)**, pp. 187-193.

[21] P. Duygulu, K. Barnard, J. F. G. de Freitas and D. A. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary", Computer Vision—ECCV 2002, Springer Berlin Heidelberg, **(2006)**, pp. 97-112.

[22] J. Shi and J. Malik, "Normalized cuts and image segmentation", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, no. 8, **(2000)**, pp. 888-905.

[23] D. Hiemstra, "Using language models for information retrieval", Taaluitgeverij Neslia Paniculata, **(2001)**.

[24] V. Lavrenko and W. B. Croft, "Relevance based language models. Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval", ACM, **(2001)**, pp. 120-127.

Authors

Ying Xia, received the Ph.D. degree in computer science and technology from the Southwest Jiaotong University, China, in 2012. Currently, she is a professor at Chongqing University of Posts and Telecommunications, China. Her research interests include spatial database, GIS, and Cross-media retrieval.

YunLong Wu, received the bachelor degree in Geographic Information System from Chongqing University of Posts and Telecommunications in 2012. Currently, he is studying for the master degree at Chongqing University of Posts and Telecommunications. His current research interests include machine learning and Cross-media retrieval.

JiangFan Feng, received his B.S. degree from Southwest Agricultural University, and his Ph.D. degree from Nanjing Normal University, in 2002 and 2007. He works as associate professor of Chongqing University of Posts and Telecommunications. His main research area includes spatial information integration and multimedia geographical information system.