HIGH TECHNOLOGY LETTERS | Vol. 20 No. 4 | Dec. 2014 | pp. 409 ~ 414

doi: 10.3772/j. issn. 1006-6748. 2014. 04. 011

Fusing PLSA model and Markov random fields for automatic image annotation①

Tian Dongping(田东平)②* **, Zhao Xiaofei**, Shi Zhongzhi**
(* Institute of Computer Software, Baoji University of Arts and Sciences, Baoji 721007, P. R. China)
(** Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, P. R. China)

Abstract

A novel image auto-annotation method is presented based on probabilistic latent semantic analysis (PLSA) model and multiple Markov random fields (MRF). A PLSA model with asymmetric modalities is first constructed to estimate the joint probability between images and semantic concepts, then a subgraph is extracted served as the corresponding structure of Markov random fields and inference over it is performed by the iterative conditional modes so as to capture the final annotation for the image. The novelty of our method mainly lies in two aspects: exploiting PLSA to estimate the joint probability between images and semantic concepts as well as multiple MRF to further explore the semantic context among keywords for accurate image annotation. To demonstrate the effectiveness of this approach, an experiment on the Corel5k dataset is conducted and its results are compared favorably with the current state-of-the-art approaches.

Key words: automatic image annotation, probabilistic latent semantic analysis (PLSA), expectation maximization, Markov random fields (MRF), image retrieval

0 Introduction

With the prevalence of digital imaging devices such as webcams, phone cameras and digital cameras, the number of accessible images is growing at an exponential speed. Thus how to make the best use of these resources becomes an emerging problem. Although content-based image retrieval (CBIR) has been studied and explored in the last decade, its performance is far from satisfactory due to the well known semantic gap. In general, people prefer to query images by semantic keywords rather than their low-level features. Alternatively, since manual annotation is expensive and difficult to be extended to large image dataset, automatic image annotation (AIA) has emerged as a striking and crucial problem in semantic based image retrieval. The state-of-the-art research on automatic image annotation has proceeded along two categories. The first one poses image annotation as a supervised classification problem[1], which treats each semantic concept as an independent class and constructs different classifiers for different concepts. This approach predicts the annotation of a new image by computing similarity at the visual level and propagating the corresponding keywords. The second category treats the words and visual tokens in each image as equivalent features in different modalities. Image annotation is then formalized by modeling the joint distribution of visual and textual features on the training data and predicting the missing textual features for a new image. As a milestone work of this perspective, Duygulu, et al.[2] proposed a translation model (TM) to treat AIA as a process of translation from a set of blob tokens to a set of keywords. Jeon, et al.[3] put forward cross-media relevance model (CMRM) to annotate image. Subsequently, CMRM is improved through continuous-space relevance model (CRM) by Manmatha, et al.[4] and multiple Bernoulli relevance model (MBRM) by Feng, et al.[5]. As a latent aspect model, PLSA has been successfully applied in multimedia processing. The method in Ref. [6] extends the PLSA by adding spatial information based on the visual words. Monay, et al.[7] proposed the representative PLSA-WORDS ap-

---

① Supported by the National Basic Research Priorities Program (No. 2013CB329502), the National High-tech R&D Program of China (No. 2012AA011003), National Natural Science Foundation of China (No. 61035003, 61072085, 60933004, 60903141) and the National Science and Technology Support Program of China (No. 2012BA107B02).
② To whom correspondence should be addressed. E-mail: tiandp@ics.ict.ac.cn
Received on Dec. 22, 2012

proach. More recently Li, et al. [8] come up with a PLSA based automatic image annotation system using two linked PLSA models to learn the mixture of aspects from both visual and textual modalities.

It is important to note that the methods aforementioned seldom address the learning of semantic context when more parameters mandatory for modeling the relationship are considered. Fortunately, Markov random fields (MRF) can incorporate various prior contextual information in a quantitative way, and it has been extensively used in computer vision applications. Qi, et al. [9] applied a MRF to video annotation. Escalante et al. [10] proposed a MRF model as part of their image annotation framework, which additionally used word-to-word correlation for the improvement of annotation systems. In addition, Hernandez-Gracidas and Sucar[11] combined several types of spatial relations under the MRF framework to improve automatic image annotation. More recent work[12] employed a MRF to model the context relationships among semantic concepts with keyword subgraphs generated from training samples for each keyword. Besides, Llorente, et al. [13] proposed a direct image retrieval framework based on MRF which exploited the semantic context dependencies of the image.

In this paper, a novel annotating model is presented by fusing PLSA and Markov random fields (PLSA-MRF) for automatic image annotation, in which multiple Markov random fields is adopted to boost the potential of the PLSA model by taking full advantage of the semantic context among keywords. The potential functions, defined by Xiang[12], are based on the PLSA model for adaptive label prediction. And the model parameters are estimated by maximum pseudo-likelihood with Gaussian prior for regularization so as to avoid the evaluation of the partition function. In addition, the PLSA-MRF model determines the number of semantic labels for an image automatically and is robust to the data imbalance problem by applying random sampling technique. Finally, the method on the Corel5k dataset is evaluated and the experimental results are competitive with several state-of-the-art approaches. The rest of the paper is organized as follows. Section 1 presents how to apply PLSA to model annotated images. In Section 2, MRF is first introduced, and then the potential functions based on the PLSA model as well as the parameter estimation of MRF are elaborated respectively specifically for the proposed PLSA-MRF image refining annotation framework. Experimental results on the Corel5k dataset are reported and analyzed in Section 3. Finally, the paper is ended with some important conclusions and future work in Section 4.

1 PLSA model

PLSA is a statistical latent class model which introduces a hidden variable (latent aspect) z_k in the generation process of each element x_j in document d_i. Given this unobservable variable z_k, each occurrence x_j is independent of the document it belongs to, which corresponds to the following joint probability:

$$P(d_i, x_j) = P(d_i) \sum_{k=1}^{K} P(z_k \mid d_i) P(x_j \mid z_k) \quad (1)$$

The model parameters of PLSA are the two conditional distributions: $P(x_j \mid z_k)$ and $P(z_k \mid d_i)$. $P(x_j \mid z_k)$ characterizes each aspect and remains valid for documents out of the training set. On the other hand, $P(z_k \mid d_i)$ is only relative to the specific documents and cannot carry any prior information to an unseen document. An EM algorithm is used to estimate the parameters through maximizing the log-likelihood of the observed data.

$$L = \sum_{i=1}^{N} \sum_{j=1}^{M} n(d_i, x_j) \log P(d_i, x_j) \quad (2)$$

where $n(d_i, x_j)$ is the number of element x_j in document d_i. The steps of the EM algorithm can be concisely described as follows.

E-step. The conditional distribution $P(z_k \mid d_i, x_j)$ is computed from the previous estimation of the parameters:

$$P(z_k \mid d_i, x_j) = \frac{P(z_k \mid d_i) P(x_j \mid z_k)}{\sum_{l=1}^{K} P(z_l \mid d_i) P(x_j \mid z_l)} \quad (3)$$

M-step. The parameters $P(x_j \mid z_k)$ and $P(z_k \mid d_i)$ are updated with the new expected values $P(z_k \mid d_i, x_j)$:

$$P(x_j \mid z_k) = \frac{\sum_{i=1}^{N} n(d_i, x_j) P(z_k \mid d_i, x_j)}{\sum_{m=1}^{M} \sum_{i=1}^{N} n(d_i, x_m) P(z_k \mid d_i, x_m)} \quad (4)$$

$$P(z_k \mid d_i) = \frac{\sum_{j=1}^{M} n(d_i, x_j) P(z_k \mid d_i, x_j)}{\sum_{j=1}^{M} n(d_i, x_j)} \quad (5)$$

If one of the parameters ($P(x_j \mid z_k)$ or $P(z_k \mid d_i)$) is known, the other one can be inferred by using fold-in method, which updates the unknown parameters with the known parameters kept fixed so that it can maximize the likelihood with respect to the previously trained parameters. Similar to Ref. [7], in this paper the joint probability between an image and the semantic concepts can be easily calculated from two linked PLSA models sharing the same distribution over aspects.

2 Multiple Markov random fields

Markov random field (MRF)[14] is a probabilistic model which combines the prior knowledge given by some observations and knowledge given by the interaction with neighbors. MRF is appealing in automatic image annotation for the main reason is that it can incorporate various prior contextual information or constraints in a quantitative way. A set of random variables $F = \{f_1, f_2, \cdots, f_m\}$ is said to be an MRF on sites $S = \{1, 2, \cdots, m\}$ with respect to a neighborhood system $N = \{N_i \mid i \in S\}$, where N_i is the set of sites neighboring i, if and only if the following two conditions are satisfied:

$$P(f) > 0, \ \forall f \in F \tag{6}$$

$$P(f_i \mid f_{S-\{i\}}) = P(f_i \mid f_{N_i}), \ \forall i \in S \tag{7}$$

where $S - \{i\}$ is the set difference, $f_{S-\{i\}}$ denotes the set of labels at the sites in $S - \{i\}$, and $f_{Ni} = \{f_{i'} \mid i' \in N_i\}$ stands for the set of labels at the sites neighboring i. According to the Hammersley-Clifford theorem, the joint probability of MRF can be expressed as

$$P(f) = Z^{-1} \times e^{-U(f)} \tag{8}$$

where $Z = \sum_f e^{-U(f)}$ is a normalizing constant called the partition function, and $U(f)$ is called the energy function. The optimal configuration can be found by minimizing the energy function $U(f)$ obtained by

$$U(f) = \sum_c V_c(f) + \lambda \sum_o V_o(f) \tag{9}$$

where $V_c(f)$ and $V_o(f)$ are potential functions. In detail, $V_c(f)$ denotes the domain information given by the neighbors and $V_o(f)$ stands for the information given by the observations. λ is a constant that weights the contribution of each term. In this paper, we only consider cliques of order up to two. Thus the energy function can be described as

$$U(f) = \sum_{i \in S} V_c(f_i) + \sum_{i \in S} \sum_{j \in N_i} V_o(f_i, f_j) \tag{10}$$

2.1 Potential functions based on the PLSA model

Keyword co-occurrence is used to define the correlations between keywords. Each training image can be considered as a "document" composed of the "associated keywords" from the predefined vocabulary set V. Given the keyword set $S = \{1, 2, \cdots, m\}$, where $i \in S$ corresponds to keyword w_i in vocabulary V, then a graph $G = (S, \varepsilon)$ on keyword set S can be constructed, where $(i, j) \in \varepsilon$ if and only if i and j are correlated. Similar to the work of Ref. [12], MRF is constructed for each keyword in the set V so as to capture different semantics among the keywords. Assume that S denotes the sites (keywords) of the single keyword MRF, the random variable $f_i \in \{-1, +1\}$ indicates the absence or presence of keyword w_i for an image. Then the site potential and edge potential functions can be defined as

$$V_c(f_i) = f_i(\lambda_i + \alpha_i P(d, w_i)) \tag{11}$$

$$V_o(f_i, f_j) = \beta_{ij} f_i f_j P(d, w_j) \tag{12}$$

where $P(d, w_i)$ denotes the joint probability of image feature d and keyword w_i calculated by the PLSA model described in Section 1, λ_i, α_i and β_{ij} are the parameters to be estimated. Hence, the energy function can be reformulated by substituting Eqs(11) and (12) into Eq. (10).

$$U(f) = \sum_{i \in S} f_i(\lambda_i + \alpha_i P(d, w_i)) + \sum_{i \in S} \sum_{j \in N_i} \beta_{ij} f_i f_j P(d, w_j) \tag{13}$$

2.2 Parameter estimation

After the joint distribution of MRF is defined, the next crucial task is to estimate its parameters. Since a probability model is incomplete if not all the parameters involved are specified, even if the functional form of the distribution is known. The widely used technique for parameter estimation in Markov random field is maximum likelihood. Here an approximation scheme is adopted called pseudo-likelihood to avoid the evaluation of the partition function, which is defined as

$$PL(f) = \prod_{i \in S} P(f_i \mid f_{N_i}) = \prod_{i \in S} \frac{e^{-U_i(f_i, f_{N_i})}}{\sum_{f_i} e^{-U_i(f_i, f_{N_i})}} \tag{14}$$

where

$$U_i(f_i, f_{N_i}) = V_c(f_i) + \sum_{j \in N_i} V_o(f_i, f_j) \tag{15}$$

Eq. (15) is the energy introduced by site i. Similarly, substituting Eqs(11) and (12) into Eq. (15), we can get:

$$U_i(f_i, f_{N_i}) = f_i(\lambda_i + \alpha_i P(d, w_i)) + \sum_{j \in N_i} \beta_{ij} f_i f_j P(d, w_j) \tag{16}$$

Let $\theta_i = (\lambda_i, \alpha_i, \beta_{ij \forall j \in N_i})^{\mathrm{T}}$ and $x_i = (1, P(d, w_i), f_j P(d, w_j)_{\forall j \in N_i})^{\mathrm{T}}$, then Eq. (16) can be rewritten as

$$U_i(f_i, f_{N_i}) = f_i \theta_i^{\mathrm{T}} x_i \tag{17}$$

where θ_i is the parameter associated with site i and x_i denotes the training data constructed for site i. Substituting Eq. (17) into Eq. (14), the pseudo-likelihood can be further expressed as

$$PL(f) = \prod_{i \in S} \frac{e^{-f_i \theta_i^{\mathrm{T}} x_i}}{e^{-\theta_i^{\mathrm{T}} x_i} + e^{\theta_i^{\mathrm{T}} x_i}} \tag{18}$$

Here parameters $\theta = (\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}}, \cdots, \theta_{|S|}^{\mathrm{T}})^{\mathrm{T}}$ can be estimated by maximizing the pseudo-likelihood with regularization on the training images by using the Newton's method. Meanwhile, iterative conditional modes for inference in the constructed MRF until convergence are employed and then the most probable labels of the sites can be obtained so as to complete the image annotation precisely. Up to this point, the procedure of PLSA-MRF for accurate image auto-annotation is presented, which is summarized in Algorithm 1.

Algorithm 1: Fusing PLSA and MRF for Automatic Image Annotation

Input: unlabeled image I, keyword vocabulary V, training set T, constructed keyword graph G
Output: labels of image I
Initialize: new training set $T_i = \Phi$ for site i, $T_w = \Phi$ for MRF$_w$, $K_i = |T_i|$ denotes the size of T_i, $|S| = m$
for each $w \in V$ **do**
  Extract a subgraph G_w from G for MRF$_w$
  for each site i of MRF_w **do**
    Sample T to get a balanced dataset T'_i
    for each $d^k \in T'_i$ **do**
      Extract labels f_i^k and f_j^k, $\forall j \in N_i$
      Calculate joint probability $P(d^k, w_i)$ and $P(d^k, w_j)$ using PLSA model
      Calculate training data $x_i^k = (1, P(d^k, w_i), f_j^k P(d^k, w_j)_{\forall j \in N_i})^{\mathrm{T}}$ for site i
    end for
    $T_i = \{(x_i^k, f_i^k)\}_{k=1}^{K_i}$
  end for
  $T_w = \cup_{i=1}^{|S|} T_i$
  Estimate the parameters of MRF$_w$ based on T_w
  Perform inference of I on MRF$_w$ to get the corresponding labels
end for

The high performance of our algorithm mainly roots in two aspects. First, the PLSA model is used to estimate the joint probability with EM algorithm rather than MBRM employed in Ref. [12], whose drawback is that estimating the joint probability of an image and its keywords requires an expectation over all training images. And the complexity of the kernel density representations may hinder applicability of MBRM to large datasets. Furthermore, MBRM requires some important parameters to be set manually, which have a significant impact on its performance. Second, the normalized cuts algorithm[15] is adopted to segment images into a number of meaningful regions instead of nonoverlapping grid for extracting image regions. Although the latter can reduce the computational complexity in comparison to using overlapping blocks or region-based segmentation approaches, the overall performance of this kind model can be quite sensitive to the block size and shifts in the image. In addition, the random sampling technique is applied to deal with the data imbalance problem as well as the early stopping technique to tackle the over-fitting problem of the PLSA model.

3 Experiments

In order to test the effectiveness and accuracy of the proposed PLSA-MRF, an experiment is conducted on the Corel5k dataset obtained from Duygulu et al.[2], which consists of 5000 images from 50 Corel Stock Photo CD. Each CD contains 100 images with a certain theme (e. g. polar bears), of which 90 are designated to be in the training set and 10 in the test set, resulting in 4500 training images and a balanced 500-image testing collection. The normalized cuts algorithm rather than nonoverlapping grid is utilized to segment 4500 images, which totally obtains 42379 segmented regions. For each image, at most 10 largest regions are selected with 36-dimensional visual features (24-dim color features and 12-dim texture features) extracted from each region. Followed by these features are clustered by k-means algorithm and discretized into clusters, which are considered as visual words. Thus, the clustering process generates a visual-word vocabulary describing different local patches in images. The number of clusters determines the size of the vocabulary. By mapping all the regions to visual words, we can represent each image as a bag of visual words. Similar to Ref. [7], the dimension for images is set to 1000 dimension in our experiment.

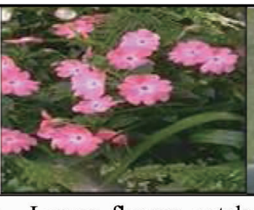
3.1 Automatic image annotation results

This paper applies MATLAB 7.0 to implement the proposed various PSO algorithms. The experiments are carried out on a 1.80 GHz Intel Core Duo CPU personal computer (PC) with 2.0G memory running Microsoft Windows XP professional. To show the effectiveness of the model (PLSA-MRF) proposed in this paper, a direct comparison with several previous approaches[2-5, 8, 12] is made. Similar to Ref. [4], recall and precision of each word in the test set are computed and the mean of these values is used to summarize its performance. The experimental results listed in Table 1 are based on two sets of words: the subset of 49 best words and the complete set of all 260 words that occur in the training set. From Table 1, it is easy to see that PLSA-MRF outperforms all the others, especially the first three approaches. Meanwhile, it is also superior to PLSA-FUSION, MBRM and MRFA[12]. In addition, Table 2 shows some annotating results (only four cases are listed here due to the limited space, and the re-ranked and new words compared to those of MRFA and the ground truth annotation are underlined and italics respectively) generated by MRFA and PLSA-MRF respectively, which further demonstrate the effectiveness of PLSA-MRF proposed in this paper.

Table 1 Performance comparison of AIA on Corel5k dataset

Models	Translation	CMRM	CRM	PLSA-FUSION	MBRM	MRFA	PLSA-MRF
#words with recall > 0	49	66	107	108	109	124	128
Results on 49 best words							
Mean per-word recall	0.34	0.48	0.70	0.76	0.68	0.67	0.78
Mean per-word Precision	0.20	0.40	0.59	0.58	0.64	0.76	0.76
Results on all 260 words							
Mean per-word recall	0.04	0.09	0.19	0.22	0.20	0.23	0.26
Mean per-word Precision	0.06	0.10	0.16	0.16	0.19	0.27	0.29

Table 2 Annotation comparison with MRFA and PLSA-MRF
(re-ranked and new words are underlined and italics respectively)

Image				
Ground Truth Annotation	Leaves, flowers, petals, stems	cars, tracks, turn, prototype	trees, people, tables, restaurant	light, shops
MRFA Annotation	Leaves, flowers, petals, stems	grass, cars, tracks, prototype	people, flowers, restaurant, shops, street, festival	light, shops
PLSA-MRF Annotation	flowers, Leaves, petals, stems	cars, tracks, grass, prototype	people, trees, tables, *festival*	light, shops, *buildings*

3.2 Ranked image retrieval results

In this section, mean average precision (mAP) is employed as a metric to evaluate the performance of single word retrieval. We only compare our model with CMRM, CRM, MBRM and PLSA-FUSION because mAP of the MRFA model cannot be accessed directly from the literature. As illustrated in Table 3, our model can obtain significant improvements of 88%, 33%, 7% and 23% mean average precision on all 260 words over CMRM, CRM, MBRM and PLSA-FUSION respectively. Correspondingly, the gains of 80%, 33%, 3% and 20% mean average precision on the set of words that have positive recalls can be achieved. In sum, the PLSA-MRF model proposed in this paper is apparently superior to other models except for MBRM with a comparatively marginal improvement.

Table 3 Ranked image retrieval results based on one word queries

Mean Average Precision for Corel Dataset		
Models	All 260 words	Words with recall > 0
CMRM	0.17	0.20
CRM	0.24	0.27
MBRM	0.30	0.35
PLSA-FUSION	0.26	0.30
PLSA-MRF	0.32	0.36

To further illustrate the effect of PLSA-MRF proposed in this paper, Fig. 1 presents the retrieval results obtained with single word queries on several challenging visual concepts being queries. Each row displays the top five matches to the semantic query "coast", "tiger", "mountain" and "flower" from top to bottom respectively. The diversity of visual appearance of the returned images bears out that our model also has good generalization ability.

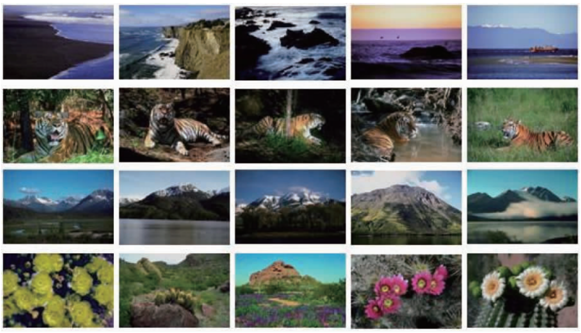


Fig. 1 Semantic retrieval results on Corel5k dataset

Finally, the complexity of PLSA-MRF proposed in this paper is analyzed. Assuming that there are D training images and each image produces R visual feature vectors, then the complexity of our model is $O(DR)$, which is similar to the classic CRM and MBRM mentioned in Ref. [1].

4 Conclusions

In this paper, a novel annotating model based on probabilistic latent semantic analysis and multiple Markov random fields is presented for precise image auto-annotation. The experimental results on the Corel5k dataset show that the model outperforms several state-of-the-art approaches. In the future, we intend to introduce semi-supervised learning into our approach for utilizing the labeled and unlabeled data simultaneously, which can alleviate the harsh requirements of a large number of labeled training images during the image annotating model construction. Meanwhile, different image datasets can be employed to detect the PLSA-MRF model proposed in this paper comprehensively.

References

[1] Carneiro G, Chan A, Moreno P, et al. Supervised learning of semantic classes for image annotation and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(3): 394-410

[2] Duygulu P, Barnard K, Freitas N De, et al. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In: Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark, 2002. 97-112

[3] Jeon L, Lavrenko V, Manmantha R. Automatic image annotation and retrieval using cross-media relevance models. In: Proceedings of the 26th International ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada, 2003. 119-126

[4] Lavrenko V, Manmatha R, Jeon J. A model for learning the semantics of pictures. In: Proceedings of the 17th International Conference on the Advances in Neural Information Processing Systems, Vancouver, Canada, 2003. 553-560

[5] Feng S, Manmatha R, Lavrenko V. Multiple Bernoulli relevance models for image and video annotation. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, Washington, USA, 2004. 1002-1009

[6] Fergus R, Li F, Perona P, et al. Learning object categories from Google's image search. In: Proceedings of the 10th International Conference on Computer Vision, Beijing, China, 2005. 1816-1823

[7] Monay F, Gatica-Perez D. Modeling semantic aspects for cross-media image indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(10): 1802-1817

[8] Li Z, Liu X, Shi Z, et al. Learning image semantics with latent aspect model. In: Proceedings of the International Conference on Multimedia and Expo, New York, USA, 2009. 366-369

[9] Qi G, Hua X, Rui Y, et al. Correlative multi-label video annotation. In: Proceedings of the 15th International Conference on Multimedia, Augsburg, Germany, 2007. 17-26

[10] Escalante H, Montes M, Sucar L. Word co-occurrence and Markov random fields for improving automatic image annotation. In: Proceedings of the 18th British Machine Vision Conference, University of Warwick, UK, 2007. 600-609

[11] Hernandez-Gracidas C, Sucar L. Markov random fields and spatial information to improve automatic image annotation. In: Proceedings of the 2nd Pacific-Rim Symposium on Image & Video Technology, Santiago, Chile, 2007. 879-892

[12] Xiang Y, Zhou X, Chua T, et al. A revisit of generative model for automatic image annotation using Markov random fields. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, Florida, USA, 2009. 1153-1160

[13] Llorente A, Manmatha R, Ruger S. Image retrieval using Markov random fields and global image features. In: Proceedings of the 9th International Conference on Image and Video Retrieval, Xi'an, China, 2010. 243-250

[14] Li S. Markov random field modeling in image analysis (3rd edition). Springer- Verlag Press, 2009

[15] Shi J, Malik J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905

Tian Dongping, born in 1981. He received his M. Sc. and Ph. D degrees from Shanghai Normal University and Institute of Computing Technology, Chinese Academy of Sciences in 2007 and 2013, respectively. His research interests include computer vision and machine learning.