Journal of Information Hiding and Multimedia Signal Processing ©2017 ISSN 2073-4212
Ubiquitous International Volume **8**, Number **1**, January **2017**

Semantic Image Annotation based on Robust Probabilistic Latent Semantic Analysis

Dongping Tian

Institute of Computer Software
Baoji University of Arts and Sciences
No.1 Hi-Tech Avenue, Hi-Tech District, Baoji, Shaanxi 721013, P.R. China

Institute of Computational Information Science
Baoji University of Arts and Sciences
No.44 Baoguang Road, Weibin District, Baoji, Shaanxi 721007, P.R. China
tiandp@ics.ict.ac.cn, tdp211@163.com

Received October, 2016; revised December, 2016

Abstract. *Automatic image annotation is a promising solution to enable the semantic image retrieval via keywords. In this paper, we present a robust probabilistic latent semantic analysis (PLSA) for the task of automatic image annotation. On the one hand, since labeled images are often hard to obtain or create in large quantities while the unlabeled ones are easier to collect. Semi-supervised learning aims to achieve good classification performance with the help of unlabelled data in the presence of the small sample size problem. Based on this recognition, the transductive support vector machine (TSVM) is exploited to enhance the quality of the training image data. On the other hand, the traditional bag-of-visual-words model is improved by integrating the contextual semantic information among visual words based on the PLSA. In the meanwhile, the approximation strategy of pseudo-likelihood in Markov random field (MRF) is introduced to combine the feature appearance similarity in feature domain and the contextual semantic information in spatial domain. Extensive experiments on the general-purpose Corel5k dataset validate that the proposed method is much more effective than several state-of-the-art approaches regarding their effectiveness and efficiency in the tasks of automatic image annotation and retrieval.*
Keywords: Automatic image annotation, PLSA, TSVM, Bag-of-visual-words, MRF

1. **Introduction.** Automatic image annotation (AIA) is a promising methodology for image retrieval. However, it is still in its infancy and is not sophisticated enough to extract perfect semantic concepts according to image low-level features, often producing noisy keywords irrelevant to image semantics that significantly hinders the task of semantic based image retrieval. Probabilistic topic model (PTM) with hidden topic variables, originally developed for statistical text modeling of large document collections, has recently become an active research topic for multi-media representation and annotation in both computer vision and pattern recognition. As a representative PTM, probabilistic latent semantic analysis (PLSA) has been successfully applied in a variety of multimedia processing tasks, including image annotation [1-12], image retrieval [13,14], image classification [15,16] and several other applications [17-24]. As for its representative applications for AIA, Monay et al. proposed a series of PLSA models for automatic image annotation [1-3], among which PLSA-MIXED [1] learned a standard PLSA model on a concatenated representation of the textual and visual features while PLSA-WORDS or PLSA-FEATURES [2,3] allowed modeling of an image as a mixture of latent aspects that was defined either by its text captions or by its visual features for which the conditional distributions over aspects were estimated from one of the two modalities only. Note that Romberg et al.[4] extended the standard single-layer probabilistic latent semantic analysis model to multiple multimodal layers that consisted of two leaf-PLSA (here from two different data modalities: image tags and visual image features) and a single top-level PLSA node

merging the two leaf-PLSA. In literature [5], a two-PLSA model was constructed for automatic image annotation. A recent work by Zheng et al.[6] developed an image annotation system with concept level search using PLSA and canonical correlation analysis (CCA), which was able to generate appropriate keywords to annotate the query images via using large-scale image database. In order to extract effective features to reflect the intrinsic content of images as complete as possible, Zhang et al.[7] proposed a multi-feature PLSA (MF-PLSA) to tackle this problem by combining low-level visual features for image region annotation in that it handled data from two different visual feature domains. In recent work of [8], Guo et al. constructed a supervised PLSA (S-PLSA) model to improve the image segmentation by using the classification results with an integrated framework based on PLSA and S-PLSA to accommodate segmentation and annotation procedures. In addition, the standard PLSA was extended to higher order for image indexing by treating images, visual features and tags as three observable variables of an aspect model [9] so as to learn a space of latent topics that incorporated the semantics of both visual and tag information. In our previous work [10], a unified two-stage refining image annotation method was presented by integrating PLSA with random walk model. Extensive experiments validated its effectiveness and efficiency. Especially in the research [11], a PLSA model with asymmetric modalities was embedded into the Markov random fields for automatic image annotation. Specifically, a PLSA model was constructed with asymmetric modalities to estimate the joint probability between an image and the semantic concepts, a subgraph served as the corresponding structure of MRF was then extracted and the inference over it was performed by the iterative conditional modes so as to capture the final annotation for the image. Alternatively, as for the improvement of PLSA model itself, it can be improved from four aspects including its initialization [25], visual words [26], hidden layers [14,27] and integration with other models [10-12]. Representative work includes the rival penalized competitive learning method [25] exploited to initialize PLSA model due to the expectation maximization is sensitive to its initialization. Li et al.[26] put forward a semantic annotation model that applied continuous PLSA and standard PLSA to deal with the visual and textual data respectively, in which an adaptive asymmetric learning approach was adopted to learn the correlation between visual and textual modalities. In [14], the standard single-layer PLSA model was further extended to the multiple multimodal layers one (MM-PLSA) for image retrieval by Romberg and Lienhart in 2012. Besides, a correlated probabilistic latent semantic analysis model [27] was proposed by introducing a correlation layer between images and latent topics to incorporate the image correlations. A recent work [28] integrated unsupervised PLSA with the k-nearest neighbor classifier for automatic landslide detection. In more recent work [12], Tian came up with a method for refining image annotation by integrating PLSA with conditional random field (CRF), whose novelty mainly lies in that exploiting PLSA to predict a candidate set of annotations with confidence scores as well as CRF to further explore the semantic context among candidate annotations for precise image annotation.

On the other hand, since labeled images are often hard to obtain or create in large quantities while the unlabeled ones are easier to collect in practical applications. semi-supervised learning (SSL) aims at learning from labeled and unlabeled data simultaneously. As for its applications in image annotation community, Zhu et al.[29] developed a semi-supervised learning based model to annotate the content of images. To be specific, the candidate annotations of unlabeled images were first obtained based on a progressive model with perceptual visual characteristics. Subsequently a random walk with restart algorithm was employed to refine these candidate annotations and the top ones were reserved as the final annotations. The work by Lu et al.[30] formulated a $L1$-norm semi-supervised learning algorithm for robust image analysis by giving new $L1$-norm formulation of Laplacian regularization that is the key step for graph-based SSL. Besides, Yuan et al.[31] proposed an approach for cross-domain image annotation by semi-supervised cross-domain learning with group sparsity, which utilized both unlabeled data in target domain and labeled data in auxiliary domain to boost the performance of AIA. Subsequent work [32] introduced the semi-supervised support vector machine into Gaussian mixture model to enable the unlabeled images to be fully exploited so as to improve its performance. In [33], Zhao et al. proposed a compact graph based semi-supervised learning method for image annotation (CGSSL) that was derived by a compact graph employed to well grasp the manifold structure. Meanwhile, an annotation system was developed in semi-supervised learning framework [34], which incorporated unnlabeled images into training phase reduced the system demand to labeled images. More recent work [35] combined multiple expert annotations using graph cuts and semi-supervised learning by considering global features and local image consistency. In addition, two semi-supervised learning algorithms, viz. self-training and co-training, were enhanced by exploring the temporal consistency of semantic concepts in video sequences for automatic video annotation [36]. Beyond this, SSL has also been used for web image interpretation [37] and k-means region clustering [38], etc. As previously reviewed, most of these approaches can achieve state-of-the-art performance and motivate us to explore better image annotation methods with the help

of their excellent experiences and knowledge. So in this paper, we present a robust PLSA model for automatic image annotation. More details of it will be described in the subsequent sections.

The rest of the paper is organized as follows. Section 2 introduces the PLSA model. In Section 3, the proposed robust probabilistic latent semantic analysis model (abbreviated as RPLSA) is elaborated from two aspects of TSVM algorithm and construction of the bag-of-visual-words (BoVW), respectively. Section 4 presents the experimental results on the Corel5k dataset. Finally, we end this paper with some concluding remarks and future work in Section 5.

2. **PLSA Model.** Probabilistic latent semantic analysis (PLSA) [39] is a statistical latent aspect model for co-occurrence data that associates an unobserved class variable with each observation, an observation being the occurrence of a word in a particular document. PLSA, in brief, is a statistical latent class model that introduces a hidden variable (latent aspect) z_k in the generative process of each element x_j in a document d_i. Given that the unobservable variable z_k, each occurrence x_j is independent of the document it belongs to, which corresponds to the following joint probability.

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

Note that Eq.(1) expresses each document as a convex combination of K aspect vectors, which amounts to the matrix decomposition. In essential, each document is modeled as a mixture of aspects – the histogram for a particular document being composed of a mixture of the histograms corresponding to each aspect. The model parameters of PLSA comprise two conditional distributions: $P(x_j|z_k)$ and $P(z_k|d_i)$, in which $P(x_j|z_k)$ characterizes each aspect and remains valid for documents out of the training set, on the other hand, $P(z_k|d_i)$ is only relative to the document-specific and cannot carry any prior information to an unseen document. Due to the existence of the sums inside the logarithm, direct maximization of the log-likelihood by partial derivatives is difficult. As a result, the expectation-maximization (EM) algorithm is usually applied to estimate the parameters through maximizing the log-likelihood function of the observed data.

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

where $n(d_i, x_j)$ denotes the number of times the term x_j occurred in document d_i. The steps of the EM algorithm can be succinctly described as follows.

E-step. The conditional distribution $P(z_k|d_i, x_j)$ is computed from the previous estimate of the parameters.

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

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

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

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

Notice that the E-step and M-step are alternated until a termination condition is met. Particularly, one can make use of a technique known as early stopping, in which one does not necessarily optimize until convergence but instead stops updating the parameters once the performance on hold-out data is not improving. This is a standard procedure that can be used to avoid the overfitting in the context of iterative fitting methods. In addition, as for the two parameters $P(x_j|z_k)$ and $P(z_k|d_i)$, if one of them is known, the other one can be inferred by using the folding-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. In the scenario of image annotation, assume that an unseen image visual features $v(d_{new})$, the conditional probability distribution $P(z_k|d_{new})$ can be inferred with the previously estimated model parameters $P(v|z_k)$. As a consequence, the posterior probabilities of each keyword can be estimated by the following equation.

$$P(w|d_{new}) = \sum_{k=1}^{K} P(w|z_k) P(z_k|d_{new}) \tag{6}$$

From Eq.(6), the top n keywords can be selected as the semantic annotations for the unseen image.

3. **The Proposed Robust PLSA.** In this section, PLSA model is first introduced. Followed by the proposed RPLSA will be elaborated from two aspects of the TSVM algorithm and bag-of-visual-words model, respectively. Fig. 1 illustrates the framework of RPLSA model proposed in this paper.

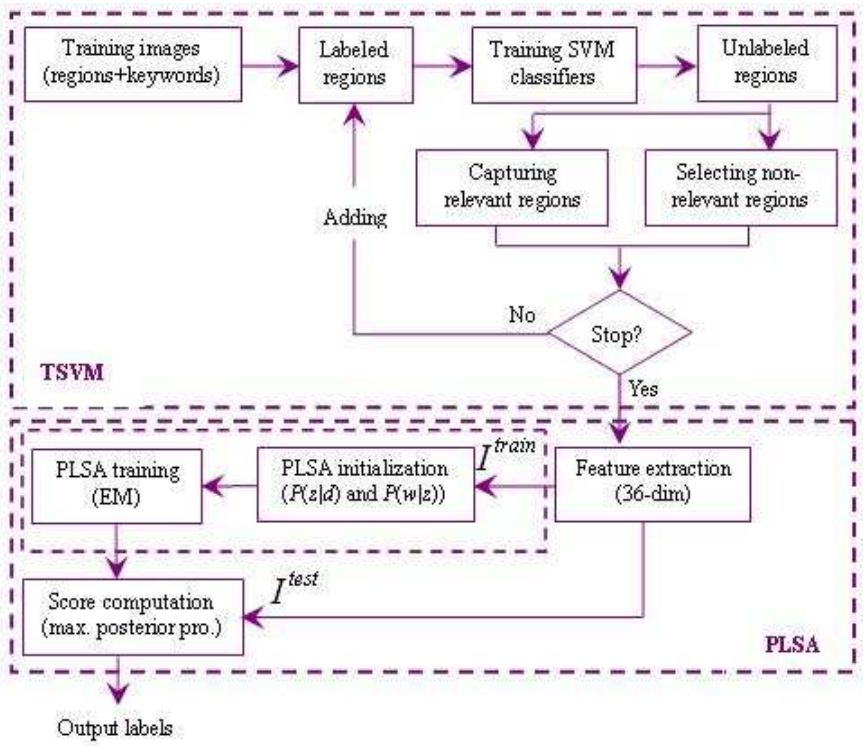

FIGURE 1. Framework of the proposed RPLSA model

3.1. **TSVM Algorithm.** Semi-supervised learning is a family of algorithms that takes advantage of both labeled and unlabeled data. Among which the transductive support vector machine is a promising way to find out the underlying relevant data from the unlabeled ones. TSVM works as follows for mining the relevant image regions: Given a keyword w, several labeled regions are taken as the relevant examples and the initial non-relevant examples are randomly sampled from the remaining regions. A two-class SVM classifier is trained first. Based on this learnt SVM classifier, the most confident relevant regions and the most non-relevant ones are added into the relevant and non-relevant training set respectively. With the expanded training set, SVM classifier will be re-trained until the maximum iteration is reached. Finally, an expanded set of labeled regions can be obtained to benefit for modeling the visual feature distribution of the keyword w. So in this paper, TSVM is adopted to explore more relevant image regions (blocks) to boost the performance of PLSA model. Its pseudocode can be describes as below.

Algorithm 1 - Pseudocode of TSVM for mining relevant regions

Input:

R_L^0 and R_U^0 denote the sets of labeled and unlabeled regions for the keyword w;
S is a SVM classifier; m, n and K denote control parameters.

Process:

for *$k = 1$ to K* ***do***
Learning a SVM classifier S from R_L^k;
Using S to classify image regions in R_U^k;
Selecting m most confidently predicted regions from R_U^k which are labeled as relevant examples;
Selecting n most confidently predicted regions from R_U^k which are labeled as non-relevant ones;
Adding $m + n$ regions with their corresponding labels into R_L^k;
Removing these $m + n$ regions from R_U^k;

Output:

R_L^k *is an expanded set of labeled image regions.*

3.2. **The Constructed BoVW.** In the recent past, many PLSA models for automatic image annotation are limited by the scope of the representation. In particular, they failed to fully exploit the contextual information of images and words. Based on this recognition and motivated by the latest research [40], a novel bag-of-visual-words model (BoVW) is constructed by integrating the contextual semantic information among visual words based on the PLSA model. Meanwhile, the approximation strategy of pseudo-likelihood in MRF is introduced to combine the feature appearance similarity in feature domain and the contextual semantic information in spatial domain. Fig. 2 illustrates the scheme of the built BoVW model. To be specific, each image is first divided into rectangular blocks in the feature domain, followed by the SIFT features of these image blocks are extracted, and the k-means algorithm is used to define visual words for image blocks, which is basically the same as the construction of the traditional bag-of-visual words model. It should be noted that, here, the Euclidean distance is utilized to measure the distance between the image blocks and visual words. On the other hand, as for the spatial correlation of image blocks, it can be estimated by the distribution of image blocks in image space. Note that both image blocks and their corresponding visual words serve as initial values of the model. Subsequently, according to Eq.(7), the contextual semantic co-occurrence relationship between the blocks and their surrounding visual words can be obtained based on the PLSA model. Finally, the Markov random field is employed to integrate the feature appearance similarity in feature domain and the contextual semantic information in spatial domain through its potential functions.

$$P(w_i|w_{N(i)}) = \frac{exp(\beta \sum_{i\in N(i)} p(w_i, w_j))}{\sum_{w_i} exp(\beta \sum_{j\in N(i)} p(w_i, w_j))} \tag{7}$$

where β is used to control the intensity of the neighborhood interaction, w_i denotes the image blocks. In addition, the distance function between image blocks and visual words is defined as below,

$$d_m^2(x_i, w_k) = \frac{d^2(x_i, w_k)}{P_G(w_i = k|w_{N(i)})} \tag{8}$$

where $P_G(w_i = k|w_{N(i)})$ denotes the prior probability of x_i belonging to class k under the conditions of neighborhood class label $w_{N(i)}$.

Up to this point, the complete procedure of the BoVW model can be succinctly described as follows.

S 1. input image blocks $X = x_i$, the maximum iteration number T, threshold ε, and the initial visual words $W = w_u$.

S 2. calculate the contextual semantic co-occurrence probability $P(w_i|w_{N(i)})$ of image blocks.

S 3. update the distance of image block and visual word. Note that if z_i denotes the corresponding visual words after image block i updated, then

$$z_i = \underset{1\leqslant k\leqslant M}{argmin}\, d_m^2(x_i, w_k) \tag{9}$$

It is also worth noting that N_i is set to 0 if z keeps invariant in two consecutive iterations, otherwise set to 1.

S 4. iterate S 3 until $max_t \left\|N^{(t)} - N^{(t+1)}\right\| < \varepsilon$ or $t > T$, then the corresponding visual words of x_i is,

$$z_i = \underset{1\leqslant k\leqslant M}{argmin}\, d_m(x_i, w_k) \tag{10}$$

else $t = t + 1$, turn to S 2.

4. **Experimental Results and Analysis.** To validate the performance of the RPLSA model, we test it on the Corel5k dataset [41], which is extensively used as basic comparative data for recent research work in image annotation. Corel5k consists of 5,000 images from 50 Corel Stock Photo CD's. 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 4,500 training images and a balanced 500-image test collection. For the sake of fair comparison, we extract similar features to [42]. That is, the images first are simply divided into a set of 32×32-sized blocks, followed by a 36-dimensional feature vector is calculated for each block, consisting of 24 color features (auto-correlogram) computed over 8 quantized colors and 3 Manhattan distances, 12 texture features (Gabor filter) computed over 3 scales and 4 orientations. As a result, each block is represented as a 36-dimensional feature vector. Finally, each image is represented as a bag of features based on the BoVW constructed in this paper.

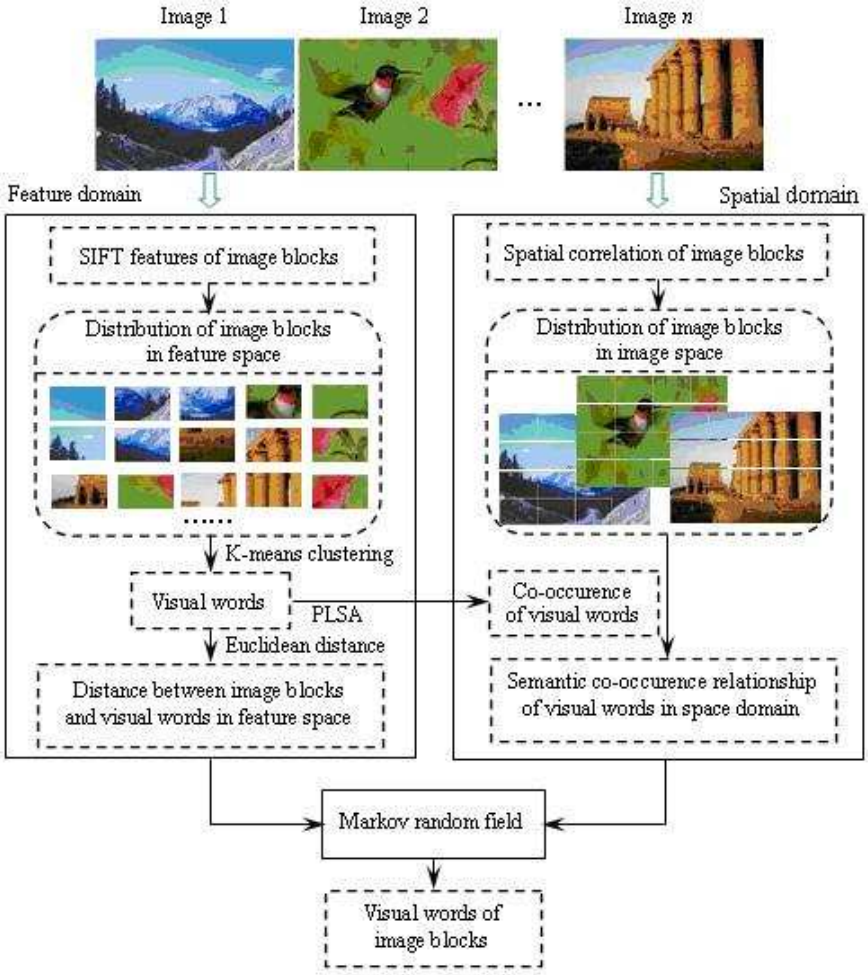

FIGURE 2. Scheme of the constructed BoVW model

To show the effectiveness of the RPLSA model proposed in this paper, we compare it with several previous approaches [3,10,11,42-45]. 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. The left part is the annotation results in which 'P' is mean precision, 'R' is mean recall and 'N+' denotes the number of keywords with non-zero recall value. The right part is the retrieval results in which the mean average precision (mAP) is used to measure the performance. Note also that "-" denotes no corresponding values can be directly acquired from the literature. From Table 1, it is clearly observed that our model markedly outperforms all the others, especially the first four approaches. In the meanwhile, it is also superior to PLSA-FUSION, MBRM, PLSA-RW and PLSA-MRF by the gains of 10, 10, 6 and 4 words with non-zero recall, 27%, 40%, 4% and 8% mean per-word recall as well as 94%, 63%, 24% and 7% mean per-word precision on the set of 260 words. This clearly validates the effectiveness of the proposed RPLSA model. Note that CRMR in Table 1 denotes the CRM with rectangular regions as input. More details on it can be gleaned from reference [42].

Fig. 3 shows some annotation results (only four cases are listed here due to the limited space) generated by PLSA-MRF and RPLSA, respectively. It can be observed that our model is able to generate more accurate annotation results compared with the original annotations as well as the ones provided in literature [11]. Taking the fourth image for example, there exist only two tags in the original annotation. However, after annotation by the RPLSA model, its annotation is enriched by the other keyword "buildings", which is very appropriate and reasonable to describe the visual content of the image. Similarly, the keyword "grass" in the second image and the "festival" in the third image.

TABLE 1. Performance comparison on Corel5k dataset

Annotation (260 words)				Retrieval (mAP)	
Models	P	R	N+	260 words	Recall>0
CMRM[43]	0.10	0.09	66	0.17	0.20
CRM[44]	0.16	0.19	107	0.24	0.27
PLSA-WORDS[3]	0.14	0.20	105	0.22	0.26
CRMR[42]	0.22	0.23	119	-	-
PLSA-FUSION[45]	0.16	0.22	122	0.26	0.30
MBRM[42]	0.19	0.20	122	0.30	0.35
PLSA-RW[10]	0.25	0.27	126	-	-
PLSA-MRF[11]	0.29	0.26	128	0.32	0.36
RPLSA	0.31	0.28	132	0.33	0.38

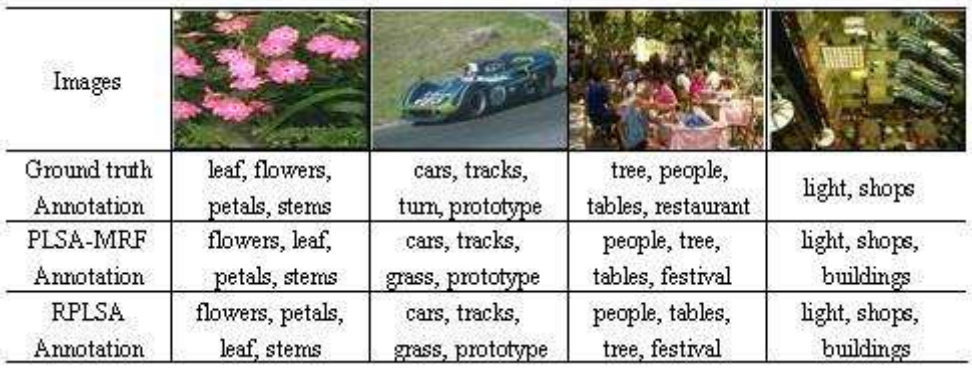

FIGURE 3. Annotation comparison with PLSA-MRF and RPLSA

To further illustrate the effect of RPLSA for automatic image annotation, Fig. 4 displays the average annotation precisions of the selected 10 words "bear", "mountain", "snow", "tree", "building", "water", "beach", "sky", "cat" and "house" based on PLSA-MRF and RPLSA, respectively. As shown in Fig. 4, the average precision of our model is consistently higher than that of PLSA-MRF.

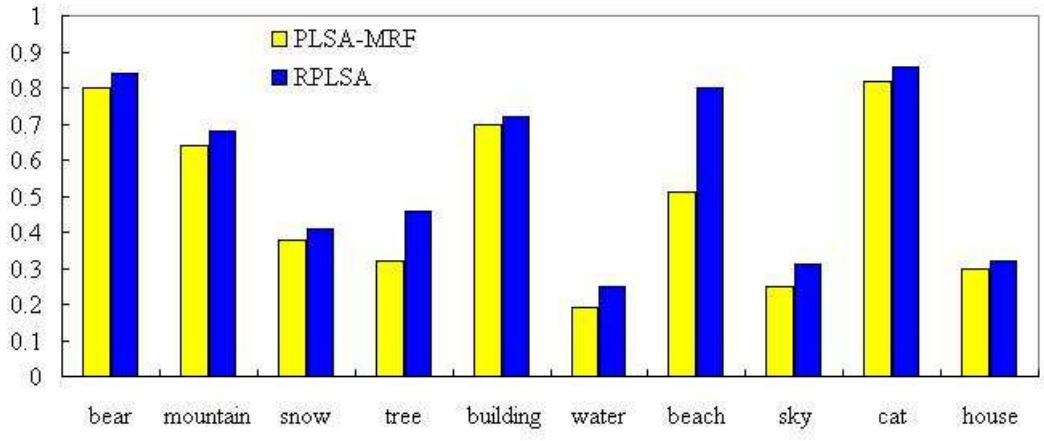

FIGURE 4. Average precision based on PLSA-MRF and RPLSA

In addition, from the perspective of probability theory, image retrieval can be seen as a procedure of ranking images in the database according to their posterior probabilities of being relevant to the query concept. To further illustrate the effect of RPLSA proposed in this paper, Fig. 5 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 "flower", "coast", "tiger" and "iceburg" from top to bottom respectively. The diversity of visual appearance of the returned images demonstrates that our model also has good generalization ability.

FIGURE 5. Semantic retrieval results on Corel5k dataset

5. **Conclusions and Future Work.** In this paper, a robust PLSA is proposed for automatic image annotation. A main novelty of this work is to formulate the bag-of-visual-words model by integrating the contextual semantic information among visual words based on PLSA as well as to combine feature appearance similarity and contextual semantic information from different domains by introducing approximation strategy of pseudo-likelihood in MRF. Extensive experiments validate its effectiveness and efficiency in the tasks of image annotation and retrieval. As for future work, we plan to delve deeper into extending this approach and applying it in wider ranges to deal with more multimedia related tasks, such as action recognition, speech recognition and other multimedia event detection tasks, etc. In addition, we plan to further explore how to integrate PLSA with other methods based on the trade-off between computational complexity and model reconstruction error for the task of automatic image annotation in the future. Last but not the least, it is worth noting that the parallelization of PLSA model to very large scale multimedia datasets is also an important issue to be further studied, especially in the current circumstances of cloud computing, cloud services, web of things, hadoop, smartwatch, fingerprint password, 3D printing and deep learning techniques, etc.

Acknowledgment. The author would like to sincerely thank the anonymous reviewers for their valuable comments and insightful suggestions that have helped us to improve the paper. Also, the author thanks Professor Zhongzhi Shi for stimulating discussions and helpful hints. This work is partially supported by the National Program on Key Basic Research Project (No.2013CB329502), the National Natural Science Foundation of China (No.61202212), the Special Research Project of the Educational Department of Shaanxi Province of China (No.15JK1038) and the Key Research Project of Baoji University of Arts and Sciences (No.ZK16047).

REFERENCES

[1] F. Monay and D. Gatica-Perez, On image auto-annotation with latent space models, *Proc. of the 11th Int'l Conf. on Multimedia (MM'03)*, pp. 275–278, 2003.

[2] F. Monay and D. Gatica-Perez, PLSA-based image auto-annotation: constraining the latent space, *Proc. of the 12th Int'l Conf. on Multimedia (MM'04)*, pp. 348–351, 2004.

[3] F. Monay and D. Gatica-Perez, Modeling semantic aspects for cross-media image indexing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1802–1817, 2007.

[4] S. Romberg, E. Horster and R. Lienhart, Multimodal PLSA on visual features and tags, *Proc. of the Int'l Conf. on Multimedia and Expo (ICME'09)*, pp. 414–417, 2009.

[5] N. Watcharapinchai, S. Aramvith and S. Siddhichai, Two-probabilistic latent semantic model for image annotation and retrieval, *Proc. of the 10th Asian Conf. on Computer Vision (ACCV'10)*, pp. 359–369, 2010.
[6] Y. Zheng, T. Takiguchi and Y. Ariki, Image annotation with concept level feature using PLSA + CCA, *Proc. of the 17th Int'l Conf. on Multimedia Modeling (MMM'11)*, pp. 454–464, 2011.
[7] R. Zhang, L. Guan, L. Zhang, et al., Multi-feature PLSA for combining visual features in image annotation, *Proc. of the 19th Int'l Conf. on Multimedia (MM'11)*, pp. 1513–1516, 2011.
[8] Q. Guo, N. Li, Y. Yang, et al., Integrating image segmentation and annotation using supervised PLSA, *Proc. of the 20th Int'l Conf. on Image Processing (ICIP'13)*, pp. 3800–3804, 2013.
[9] S. Nikolopoulos, S. Zafeiriou, I. Patras, et al., High order PLSA for indexing tagged images, *Signal Processing*, vol. 93, no. 8, pp. 2212–2228, 2013.
[10] D. Tian, X. Zhao and Z. Shi, An efficient refining image annotation technique by combining probabilistic latent semantic analysis and random walk model, *Intelligent Automation & Soft Computing*, vol. 20, no. 3, pp. 335–345, 2014.
[11] D. Tian, X. Zhao and Z. Shi, Fusing PLSA model and Markov random fields for automatic image annotation, *High Technology Letters*, vol. 20, no. 4, pp. 409–414, 2014.
[12] D. Tian, Exploiting PLSA model and conditional random field for refining image annotation, *High Technology Letters*, vol. 21, no. 1, pp. 78–84, 2015.
[13] I. Sayad, J. Martinet, T. Urruty, et al., Toward a higher-level visual representation for content-based image retrieval, *Multimedia Tools and Applications*, vol. 60, no. 2, pp. 455–482, 2012.
[14] S. Romberg and R. Lienhart, Multimodal image retrieval: fusing modalities with multilayer multimodal PLSA, *International Journal of Multimedia Information Retrieval*, vol. 1, no. 1, pp. 31–44, 2012.
[15] B. Jin, W. Hu and H. Wang, Image classification based on PLSA fusing spatial relationships between topics, *IEEE Signal Processing Letters*, vol. 19, no. 3, pp. 151–154, 2012.
[16] Y. Jiang, J. Liu, Z. Li, et al., Co-regularized PLSA for multi-view clustering, *Proc. of the 11th Asian Conf. on Computer Vision (ACCV'12)*, pp. 202–213, 2012.
[17] Y. Zhou and J. Luo, Geo-location inference on news articles via multimodal PLSA, *Proc. of the 20th Int'l Conf. on Multimedia (MM'12)*, pp. 741–744, 2012.
[18] J. Wang, P. Liu, M. She, et al., Supervised learning probabilistic latent semantic analysis for human motion analysis, *Neurocomputing*, vol. 100, pp. 134–143, 2013.
[19] Y. Ye, S. Gong, C. Liu, et al., Online belief propagation algorithm for probabilistic latent semantic analysis, *Frontiers of Computer Science*, vol. 7, no. 4, pp. 526–535, 2013.
[20] J. Wang, X. Sun, S. Nahavandi, et al., Multichannel biomedical time series clustering via hierarchical probabilistic latent semantic analysis, *Computer Methods and Programs in Biomedicine*, vol. 117, no. 2, pp. 238–246, 2014.
[21] X. Li, Q Lv and W. Huang, Learning similarity with probabilistic latent semantic analysis for image retrieval, *Ksii Transactions on Internet & Information Systems*, vol. 9, no. 4, pp. 1424–1440, 2015.
[22] Y. Zhong, Q. Zhu and L. Zhang, Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery, *IEEE Transactions on Geoscience & Remote Sensing*, vol. 53, no. 11, pp. 1–16, 2015.
[23] R. Fernandez-Beltran and F. Pla, Incremental probabilistic latent semantic analysis for video retrieval, *Image and Vision Computing*, vol. 38, pp. 1–12, 2015.
[24] X. Yang, Q. Sun and T. Wang, Blind image quality assessment via probabilistic latent semantic analysis, *SpringerPlus*, vol. 5, no. 1, pp. 1714–1731, 2016.
[25] Z. Lu, Y. Peng and H. Horace, Image categorization via robust PLSA, *Pattern Recognition Letters*, vol. 31, no. 1, pp. 36–43, 2010.
[26] Z. Li, Z. Shi, X. Liu, et al., Modeling continuous visual features for semantic image annotation and retrieval, *Pattern Recognition Letters*, vol. 32, no. 3, pp. 516–523, 2011.
[27] P. Li, J. Cheng, Z. Li, et al., Correlated PLSA for image clustering, *Proc. of the 17th Int'l Conf. on Multimedia Modeling (MMM'11)*, pp. 307–316, 2011.
[28] G. Cheng, L. Guo, T. Zhao, et al., Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and PLSA, *International Journal of Remote Sensing*, vol. 34, no. 1, pp. 45–59, 2013.
[29] S. Zhu and Y. Liu, Semi-supervised learning model based efficient Image annotation, *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 989–992, 2009.
[30] Z. Lu and Y. Peng, Robust image analysis by $L1$-norm semi-supervised learning, *In arXiv 1110.3109v1*, pp. 1–11, 2011.

[31] Y. Yuan, F. Wu, J. Shao, et al., Image annotation by semi-supervised cross-domain learning with group sparsity, *Journal of Visual Communication and Image Representation*, vol. 24, no. 2, pp. 95–102, 2013.
[32] D. Tian, Semi-supervised learning for refining image annotation based on random walk model, *Knowledge-Based Systems*, vol. 72, no. 12, pp. 72–80, 2014.
[33] M. Zhao, T. Chow, Z. Zhang, et al., Automatic image annotation via compact graph based semi-supervised learning, *Knowledge-Based Systems*, vol. 76, no. 3, pp. 148–165, 2015.
[34] S. Amiri and M. Jamzad, Automatic image annotation using semi-supervised generative modeling, *Pattern Recognition*, vol. 48, no. 1, pp. 174–188, 2015.
[35] D. Mahapatra, Combining multiple expert annotations using semi-supervised learning and graph cuts for medical image segmentation, *Computer Vision and Image Understanding*, vol. 151, pp. 114–123, 2016.
[36] M. Wang, X. Hua, L. Dai, et al., Enhanced semi-supervised learning for automatic video annotation, *Proc. of the Int'l Conf. on Multimedia and Expo (ICME'06)*, pp. 1485–1488, 2006.
[37] F. Wu, D. Xia, Y. Zhuang, et al., Web image interpretation: semi-supervised mining annotated words, *Proc. of the Int'l Conf. on Multimedia and Expo (ICME'09)*, pp. 1512–1515, 2009.
[38] A. Sayar and F. Vural, Image annotation with semi-supervised clustering, *Proc. of the 24th Int'l Symposium on Computer and Information Sciences (ISCIS'09)*, pp. 12–17, 2009.
[39] T. Hofmann, Unsupervised learning by probabilistic latent semantic analysis, *Machine Learning*, vol. 42, no. 1, pp. 177–196, 2001.
[40] Y. Liu, D. Xu, S. Feng, et al., A novel visual words definition algorithm of image patch based on contextual semantic information, *Acta Electronic Sinica*, vol. 38, no. 5, pp. 1156–1161, 2010.
[41] P. Duygulu, K. Barnard, N. de Freitas, et al., Object recognition as machine translation: learning a lexicon for a fixed image vocabulary, *Proc. of the European Conf. on Computer Vision (ECCV'02)*, pp. 97–112, 2002.
[42] S. Feng, R. Manmatha and V. Lavrenko, Multiple Bernoulli relevance models for image and video annotation, *Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition (CVPR'04)*, pp. 1002–1009, 2004.
[43] L. Jeon, V. Lavrenko and R. Manmantha, Automatic image annotation and retrieval using cross-media relevance model, *Proc. of the 26th Int'l Conf. on Research and Development in Information Retrieval (SIGIR'03)*, pp. 119–126, 2003.
[44] V. Lavrenko, R. Manmatha and J. Jeon, A model for learning the semantics of pictures, *Advances in Neural Information Processing Systems 16 (NIPS'03)*, pp. 553–560, 2003.
[45] Z. Li, Z. Shi, X. Liu, et al., Fusing semantic aspects for image annotation and retrieval, *Journal of Visual Communication and Image Representation*, vol. 21, no. 8, pp. 798–805, 2010.