Automatic Image Annotation with Cooperation of Concept-Specific and Universal Visual Vocabularies

Yanjie Wang, Xiabi Liu, and Yunde Jia

Beijing Laboratory of Intelligent Information Technology,
School of Computer Science, Beijing Institute of Technology, Beijing, P.R. China
{wangyanjie,liuxiabi,jiayunde}@bit.edu.cn

Abstract. This paper proposes an automatic image annotation method based on concept-specific image representation and discriminative learning. Firstly, the concept-specific visual vocabularies are generated by assuming that localized features from the images with a specific concept are of the distribution of Gaussian Mixture Model (GMM). Each component in the GMM is taken as a visual token of the concept. The visual tokens of all the concepts are clustered to obtain a universal token set. Secondly, the image is represented as a concept-specific feature vector by computing the average posterior probabilities of being each universal visual token for all the localized features and assigning it to corresponding concept-specific visual tokens. Thus the feature vector for an image varies with different concepts. Finally, we implement image annotation and retrieval under a discriminative learning framework of Bayesian classifiers, Max-Min posterior Pseudo-probabilities (MMP). The proposed method were evaluated on the popular Corel-5K database. The experimental results with comparisons to state-of-the-art show that our method is promising.

Keywords: Image annotation, Image retrieval, Visual vocabulary, Bag-of-features, Max-Min posterior Pseudo-probabilities (MMP).

1 Introduction

Many users of image retrieval systems prefer keyword query to visual query. Therefore, with the rapid growth of the number of available images, it becomes more and more important to automatically annotate images with keywords. An increasing interest in solving this problem of automatic image annotation (AIA) has been shown in recent literature. The AIA problem can be described as the one of associating the image descriptor with its concepts. The direct descriptor of an image is the visual features from it. Recently, the researchers also try to explore indirect textural cues, such as the text around the image in the web page, to supplement the visual features [17].

This paper investigates the problem of determining the concepts of an image according to localized visual features from it. A corresponding AIA method is proposed based on a novel concept-specific image representation schema and

S. Boll et al. (Eds.): MMM 2010, LNCS 5916, pp. 262–272, 2010.

a discriminative learning framework of Bayesian classifier, Max-Min posterior Pseudo-probabilities (MMP) [12]. Firstly, a concept-specific visual vocabulary is generated for each concept by assuming that localized features from the images with this concept are of the distribution of Gaussian Mixture Model (GMM). The GMM for each concept is learned from the training images with this concept by using the Expectation-Maximization (EM) algorithm with the Minimum Description Length criterion (MDL). Each component in the GMM is taken as a visual token of this concept. Secondly, the Concept-specific Visual Tokens (CVTs) of all the concepts are clustered to obtain a universal visual vocabulary. The correspondences between concept-specific and universal visual tokens are recorded. Thirdly, an image is represented through the cooperation between concept-specific and universal visual vocabularies. In fact, the posterior probabilities of being Universal Visual Tokens (UVTs) for localized features in the image are computed. Then the average posterior probability is calculated for each UVT and assigned to the corresponding CVT. This probability can be seen as the possibility of a CVT occurring in the image. Therefore, the average posterior probabilities for all the CVTs of each concept are arranged orderly to represent the image. It means that the feature vector for an image varies with different concepts. Finally, feature vectors extracted from images with a specific concept are also assumed to be of the distribution of GMM. By embedding this GMM into the discriminative learning framework of MMP, we get our image annotation and retrieval algorithm and evaluate it on the Corel-5K database. In the experiments, the localized feature of the image is obtained by partitioning an image into fix-sited rectangular blocks and extract Discrete Cosine Transform (DCT) features in YBR color space from each block. The performance comparison between our method and other state-of-the-art counterparts shows that the proposed method is promising. The main contributions of this work are:

(1) A novel image representation is realized for AIA through the cooperation between concept-specific and universal visual vocabularies. This image representation brings the advantages of more tolerance to background clutter, free of dimensionality reduction, and flexibility to the change of concept set. The last advantage seems useful for real applications where the number of concepts is not easy to be preset beforehand.

(2) A new discriminative learning framework of Bayesian classifier, Max-Min posterior Pseudo-probabilities (MMP), is tailored to tackle the AIA problem. Because the dimensionality of the feature vectors of an image varies with different concepts, the corresponding statistical models of the concepts learned by traditional generative learning methods such as EM algorithm are not appropriate for image classification. As a discriminative learning approach, MMP is a suitable solution to this problem.

The rest of this paper is organized as follows. Section 2 reviews the related work of AIA based on statistical modeling and learning of concepts. Section 3 presents the generation method of concept-specific and universal visual vocabularies. The cooperative image representation strategy and subsequent image annotation by

MMP is put forward in Section 4. Section 5 discusses the experimental results on Corel-5K database. We conclude the paper in Section 6.

2 Related Work

Many AIA algorithms have been developed based on statistical modeling and learning of concepts. Bayes theory is the basis of these algorithms, where the posterior probabilities of being keywords for images are used as annotation confidence but mostly reflected by joint probabilities of keywords and images or class-conditional probabilities of images given keywords. An image is usually a combination of several concepts. The regional features of an image should be expressed to realize multi-label annotation for images with multiple concepts. This task can be implemented by segmenting an image into regions or partitioning it into blocks (also called grids). Based on the localized features from segmented regions or partitioned blocks, we can directly associate the regions with concepts, or holistically classify the image according to combined regional features. Many statistical models have been explored in AIA algorithms with segmentation or partition strategy, including the translation model [5], Cross-Media Relevance Model (CMRM) [8], Continuous-space Relevance Model (CRM) [10], Coherent Language Model (CLM) [9], CORRespondence Latent Dirichlet Dllocation (CORR-LDA) [1], Hidden Concept Model [19], GMM [2], probabilistic latent semantic analysis [13], Bayes method [2], HMM [11], Bayes Point Machines (BPM) [3], etc.

Visual tokens based image representation was used behind some of AIA methods mentioned above, such as translation model, CMRM, CLM, etc. In these methods, image regions are clustered to obtain the basic elements for representing images. These basic elements are called blob-token, visual term, visual word, or visual token. In the field of object categorization which is closely related to AIA, visual tokens based image representation has also recently become a hot topic. However, most of visual token based image representation methods construct visual vocabulary by unsupervised manner, without taking class information into account. Recently, statistical modeling of visual tokens has been advised to improve its effectiveness for image classification [6,18,15]. The relation between local features and visual tokens can be described more accurately and reliably through statistical modeling of visual tokens. Furthermore, a local feature is allowed to be softly mapped to multiple visual tokens in this way, so the aliasing effects can be reduced.

3 Visual Vocabularies Generation

This section describes our method of generating concept-specific visual tokens (CVTs) and universal visual tokens (UVTs) from localized features. As illustrated in Fig. 1, we firstly generate CVTs of each concept according to localized features which are extracted from images with this concept, and then perform

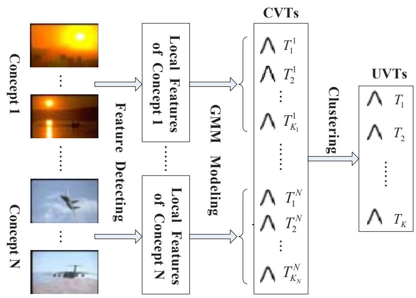

Fig. 1. The flowchart of concept-specific and universal visual vocabularies generation

the clustering on CVTs from all the concepts to obtain UVTs. The correspondences between CVTs and UVTs are recorded for subsequent use in the image representation.

3.1 Concept-Specific Visual Vocabularies Generation

In order to generate CVTs, it is assumed that the distribution of localized features extracted from images with a specific concept is a GMM. Let $\mathbf{x}$ be a localized feature, K_i be the number of components in the GMM for the i-th concept, Θ be the parameter set of the GMM for the i-th concept, which includes the weights $w_k|_{k=1}^{K_i}$, the means $\boldsymbol{\mu}_k|_{k=1}^{K_i}$, and the covariance matrices $\boldsymbol{\Sigma}_k|_{k=1}^{K_i}$. Then we have

$$p(\mathbf{x}|\Theta) = \sum_{k=1}^{K_i} w_k N(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{1}$$

where

$$\begin{aligned} &N(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\ &= (2\pi)^{-\frac{D}{2}} |\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}-\boldsymbol{\mu}_k)\right). \end{aligned} \tag{2}$$

The covariance matrix $\boldsymbol{\Sigma}_k$ is considered as a diagonal matrix for simplicity.

The parameter set Θ of the GMM is estimated by Expectation-Maximization algorithm [4] with maximum likelihood setting, which is implemented using Torch machine learning library [1] in this paper. The component number of the GMM K_i is determined by the Minimum Description Length (MDL) principle [7]. After the GMM for a concept is learned from the data by MDL-EM algorithm, each Gaussian component in the GMM is regarded as a CVT. All the Gaussian components constitute a set of CVTs for this semantic concept. Let T_j^i be the j-th CVT of the i-th CVT, then the set of CVTs for the concept is denoted as $\{T_1^i, T_2^i, \cdots, T_{K_i}^i\}$.

[1] Torch Machine Learning Library. Available: http://www.torch.ch

3.2 Universal Visual Vocabulary Generation

After the CVTs of all the concepts are generated, the k-means clustering algorithm is performed on mean vectors of the GMMs for all the CVTs to combine similar CVTs. Each resultant cluster corresponds to a universal visual token (UVT), which is composed of one or several CVTs. We get local features corresponding with CVTs in each cluster and then compute the mean vector and covariance matrix of these local features to form a UVT. Let T_i be the i-th UVT, K be the number of UVTs, then the set of UVTs is $\{T_1, T_2, \cdots, T_K\}$.

The correspondences between UVTs and CVTs are recorded, which will be used in subsequent image representation procedure. Fig. 2 illustrates the correspondences between UVTs and CVTs.

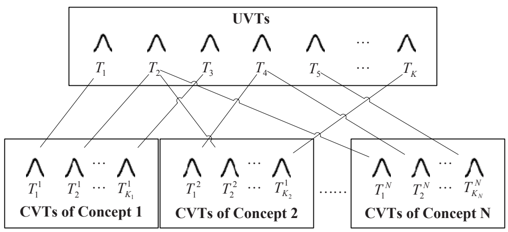

Fig. 2. Illustration of correspondences between CVTs and UVTs

4 Image Representation, Annotation and Retrieval

This section explains how to represent images through the cooperation between class-specific and universal visual vocabularies. Then the image annotation method by applying the proposed image representation scheme under GMM-MMP classification framework is described.

4.1 Cooperative Image Representation

An image is represented as a soft histogram over CVTs of each concept. The value in each bin of the histogram represents the possibility of a CVT occurring in the image, but it is measured according to the set of UVTs. Actually, the posterior probabilities of being each UVT for localized features in the image are computed using Bayes formula. Then the average posterior probability is calculated for each UVT as the measure of its occurrence possibility and assigned to the corresponding CVTs which are classified into the cluster represented by this UVT. Finally, the average probabilities for all the CVTs of a concept is arranged orderly to obtain concept-specific feature vector of the image. It means the feature vector of the image varies with different concepts. The flowchart of our image representation strategy described above is shown in Fig. 3, where the average posterior probability for a UVT and its corresponding CVTs are displayed in a same color.

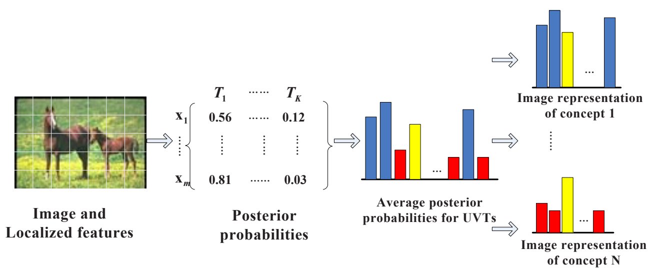

Fig. 3. Flowchart of image representation for different concepts

The details of our image representation scheme are given as follows. Firstly, since the distribution of a UVT T_i is a Gaussian model, the class-conditional probability of a localized feature $\mathbf{x}$ give T_i is computed as

$$p(\mathbf{x}|T_i) = N(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \tag{3}$$

According to Bayes formula and the assumption of the same $P(T_i)$ for all the UVTs, the posterior probability of being T_i for $\mathbf{x}$ is obtained as

$$P(T_i|\mathbf{x}) = \frac{P(\mathbf{x}|T_i)}{\sum_{k=1}^{K} P(\mathbf{x}|T_k)}. \tag{4}$$

The assignment of average posterior probabilities for UVTs to corresponding CVTs is illustrated in Fig. 3, where the red and blue columns represent the average posterior probabilities for CVTs of these two concepts and assigned from different UVTs, respectively. While the yellow columns represent the average posterior probabilities for CVTs of these two concepts but assigned from a common UVT.

The concept-specific image representation described above is more compact and more discriminative than traditional histograms over universal visual vocabulary. In traditional histograms over universal visual vocabulary, localized features of an image are mapped to thousands of the UVTs. It leads to a sparse representation. Oppositely, we obtain a compact representation by only considering the CVTs related to a specific concept. The dimensionality of feature vectors is reduced greatly. Furthermore, the occurrence possibilities of CVTs are expected to be measured as high values for images with the corresponding concept and measured as low values for images without corresponding concept. It means that our concept-specific image representation could have better discriminability than traditional histograms over universal visual vocabulary.

4.2 GMM-MMP Based Image Annotation and Retrieval

MMP is a new kind of discriminative learning approach for Bayesian classifiers. In the following, we briefly introduce the MMP algorithm designed for image annotation and retrieval. The reader is referred to our paper for more details of MMP [12].

4.2.1 Image Annotation and Retrieval by Posterior Pseudo-probabilities

Let $\mathbf{X}^C$ be a concept-specific feature vector of a concept C, which is extracted from an arbitrary image. Let $p(\mathbf{X}^C|C)$ be the class-conditional probability density function. Then the posterior pseudo-probability of being C for $\mathbf{X}^C$ is computed as

$$f(p(\mathbf{X}^C|C)) = 1 - \exp(-\lambda p^t(\mathbf{X}^C|C)), \tag{5}$$

where λ, t are positive numbers. Consequently, $f(p(\mathbf{X}^C|C))$ is a smooth, monotonically increasing function of $p(\mathbf{X}^C|C)$, and $f(0) = 0$ and $f(+\infty) = 1$. The form of class-conditional probability density function $p(\mathbf{X}^C|C)$ in Eq. 5 should be provided for using posterior pseudo-probabilities based classifers, which is also assumed to be the GMM with diagonal covariance matrix in this paper.

Given an input image, we compute the posterior pseudo-probability for each concept according to Eq. 5. Then the image is annotated through ranking concepts in descending order of their posterior pseudo-probabilities. The corresponding semantic retrieval is realized by ranking images for query concept. Given a query concept, we retrieval the images by ranking the images in descending order of the posterior pseudo-probabilities for this concept and each image in the database, which have been computed in the image annotation stage.

4.2.2 MMP Training

There are unknown parameters in Eq. 5, including λ, t, and those in $p(\mathbf{X}^C|C)$. A method called Max-Min posterior Pseudo-probabilities (MMP) is used to learn these parameters. The main idea behind MMP learning is to optimize the classifier performance through maximizing posterior pseudo-probabilities towards 1 for each class and its positive samples, while minimizing those towards 0 for each class and its negative samples. More formally, let $f(\mathbf{X}; \tilde{\mathbf{\Lambda}})$ be the posterior pseudo-probability measure function of a class, where $\tilde{\mathbf{\Lambda}}$ denote the set of unknown parameters in it. Let $\hat{\mathbf{X}}_i$ be the feature vector of arbitrary positive sample of the concept, $\bar{\mathbf{X}}_i$ the feature vector of arbitrary negative sample of the concept, m and n be the number of positive and negative samples of the concept, respectively. According to the idea above of the MMP learning, the objective function for estimating parameters is designed as

$$F(\tilde{\mathbf{\Lambda}}) = \frac{1}{m}\sum_{i=1}^{m}[f(\hat{\mathbf{X}}_i; \tilde{\mathbf{\Lambda}}) - 1]^2 + \frac{1}{n}\sum_{i=1}^{n}[f(\bar{\mathbf{X}}_i; \tilde{\mathbf{\Lambda}})]^2. \tag{6}$$

$F(\tilde{\mathbf{\Lambda}}) = 0$ means the perfect classification performance on the training data. Consequently, we can obtain the optimum parameter set $\tilde{\mathbf{\Lambda}}^*$ of the posterior pseudo-probability measure function by minimizing $F(\tilde{\mathbf{\Lambda}})$:

$$\tilde{\mathbf{\Lambda}}^* = \arg\min_{\tilde{\mathbf{\Lambda}}} F(\tilde{\mathbf{\Lambda}}). \tag{7}$$

The gradient descent algorithm is employed to optimize the parameter set $\tilde{\mathbf{\Lambda}}^*$.

5 Experiments

The image annotation and retrieval experiments were conducted on the popular Corel-5K database [5]. There are 5000 images from 50 Stock Photo CDs in this database, and each CD contains 100 digital photos of the same topic. One to five keywords are provided for each of these images. Following the commonly used evaluation scheme on Corel-5K database [5], we used 4500 images as the training set and the remaining 500 images as the test set. The total 371 semantic concepts are involved in the database, but only 260 concepts coexist in both of the training set and the test set. These 260 concepts are considered in the experiments.

We obtain localized features by partitioning an image into fix-sized rectangular blocks and extract Discrete Cosine Transform (DCT) features in YBR color space from each block. This localized feature extractor is similar with that used by Carneiro et al. [2]. The size of rectangular blocks is set to be 8×8 through experiments.

In the MMP training, the positive samples of each concept are images with the concept, and other images are its negative samples. Before using MMP training algorithm, we obtain the initial parameters by using the MDL-EM algorithm described in Section 3 on positive samples to get the parameters in the GMM, and set λ and t through experiments. The initial parameters are then revised by performing MMP training on all the samples including positive samples and negative samples. For the MDL based model selection of the GMM, we evaluate the component numbers from 20 to 300 at intervals of 10 for concept-specific visual vocabularies, and the component numbers from 1 to 20 for concepts. The resultant component numbers for concept-specific visual vocabularies vary from 20 to 280, while those for concepts are 6 to 15. As for the number of universal visual tokens, we set it to 3000 through careful experiments.

In order to compare our method with other related work on Corel-5K database, we annotate the images with top-five concepts. Then the image annotation performance is evaluated by the mean recall rate and precision rate, as well as the number of concepts with nonzero recall rate. And the subsequent image retrieval performance is evaluated by the mean average precision (MAP) [2]. Fig. 4 and Fig. 5 show the effectiveness of our image retrieval and annotation algorithm through some example. In Fig. 4, the concepts automatically annotated by the proposed method are compared with the ground-truth of human annotation for the test set. In Fig. 5, top-5 images retrieved for some query concepts are displayed from left to right.

The performance of our image annotation algorithm is compared with those recently reported on the Corel-5K database, including the co-occurrence model [14], the translation model [5], the continuous-space relevance model [10,16], the multiple-Bernoulli relevance model (MBRM) [16], and supervised multiclass labeling model (SML)[2]. We further compare our retrieval results with those from SML and MBRM. The comparison result of image annotation and retrieval is listed in Table 1-2, respectively. As shown in Table 1, our proposed algorithm achieves 26.8% recall rate and 23.5% precision rate, which are comparable to

Images				
Truth	jet, plane, sky, smoke	bengal, cat, forest, tiger	frost, frozen, ice	filed, foals, horses, mare
Our Results	smoke, plane, jet, formation, sand	tiger, cat, bengal, forest, river	crystals, frost, frozen, ice, sculpture	horses, mare, porcupine, foals, field
Images				
Truth	people, pool, swimmers, water	locomotive, railroad, smoke, train	bear, polar, snow	cars, formula, tracks, wall
Our Results	swimmers, pool, people, water, boats	train, railroad,smoke, locomotive,monastery	bear, polar, snow, herd, bush	formula, car, tracks, wall, street

Fig. 4. Comparison of the annotations by the proposed algorithm with the ground-truth

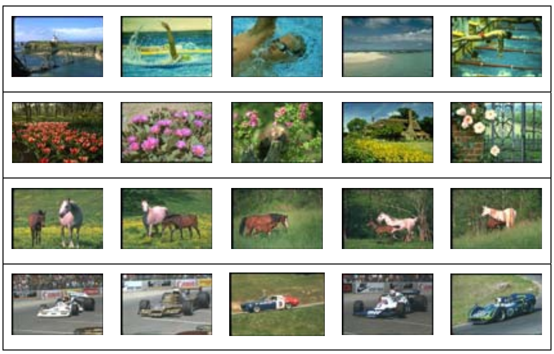

Fig. 5. Each row shows the top five retrieved images for a semantic concept. From top to bottom: water, flower, horses, and cars.

the previous best results achieved by SML and MBRM, and outperform others in both recall rate and precision rate. The number of recalled concepts is 137, which is the same as the previous best one reported by SML. As shown in Table 2, our mean average precision over the total 260 concepts is a little worse than those from SML and MBRM, while the result over concepts with nonzero recall rate is in the middle of those from SML and MBRM. It should be noted that the dense sampling features are adopted in SML. Compared overlapping blocks used there, the number of non-overlapping blocks in this work is much smaller. We expect to further show the advantages of our method by analyzing its efficiency and effectiveness according to more experiments on localized features from dense sampling in the future.

Table 1. Comparison of Automatic Annotation on Corel-5K

Methods	Co-occurrence	Translation	CRM	MBRM	SML	Our
#concepts with recall > 0	19	49	107	122	137	137
Results on all 260 words						
Mean Per-concept Recall Rate	0.02	0.04	0.19	0.25	0.29	0.268
Mean Per-concept Precision Rate	0.03	0.06	0.16	0.24	0.23	0.235

Table 2. Comparison of Semantic Retrieval on Corel-5K

Mean Average Precision for Corel-5K		
Methods	All 260 Concepts	Concepts with Recall>0
Ours	0.286	0.475
MBRM	0.30	0.35
SML	0.31	0.49

6 Conclusions

In this paper, a novel image annotation method has been proposed through representing images based on the cooperation between concept-specific and universal visual vocabulary. The main feature of the proposed method is that the image representation is defined on concept level, instead of on universal level. And the feature vector of the image varies with different concepts. For each concept, the posterior probabilities for concept-specific visual tokens and localized features in the image are measured according to the universal visual vocabulary. They are arranged orderly to represent the image. The advantages of this representation strategy are summarized as follows: 1) The image is represented with only concept-specific information to obtain more robustness to background clutter; 2) The dimensionality of feature vector of the image is small. So it is unnecessary to perform dimensionality reduction which risks the loss of discriminative information; 3) Clustering is performed on the concept-specific visual tokens, instead of on the huge set of localized features. Thus the visual token generation is more flexible to the change of concept sets. This feature seems useful for real applications where the number of concepts is not easy to be preset beforehand.

By embedding our image representation scheme into a new discriminative learning framework of Bayesian classifiers, Max-Min posterior Pseudo-probabilities (MMP), we get the corresponding image annotation and retrieval algorithm which achieved the comparable performance to the previous best methods on the Corel-5K database.

Our future work includes: 1) More sophisticated localized features such as dense sampling features will be considered to improve the effectiveness of the proposed method; 2) The proposed method will be evaluated on other more complicated databases, such as Corel-30K, PSU, etc.

Acknowledgements

This work was partially supported by National Natural Science Foundation of China (Grant No. 60973059 and Grant No. 90920009).

References

1. Blei, D.M., Jordan, M.I.: Modeling annotated data. In: ACM SIGIR Conference (2003)
2. Carneiro, G., Chan, A., Moreno, P., Vasconcelos, N.: Supervised learning of semantic classes for image annotation and retrieval. IEEE Transaction on Pattern Analysis and Machine intelligence 29(3), 394–410 (2007)
3. Chang, E., Goh, K., Sychay, G., Wu, G.: CBSA: Content-based soft annotation for multimodal image retrieval using bayes point machines. IEEE Transactions on Circuits and Systems for Video Technology 13(1), 26–38 (2003)
4. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 39(1), 1–38 (1977)
5. Duygulu, P., Barnard, K., de Freitas, J.F.G., Forsyth, D.: Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 97–112. Springer, Heidelberg (2002)
6. Farquhar, J., Szedmak, S., Meng, H., Shawe-Taylor, J.: Improving "bag-of-keypoints" image categorisation: Generative models and pdf-kernels. Technical report, University of Southampton (2005)
7. Hansen, M.H., Yu, B.: Model selection and the principle of minimum description length. Journal of American Statistical Association 96(454), 746–774 (2001)
8. Jeon, J., Lavrenko, V., Manmatha, R.: Automatic image annotation and retrieval using cross-media relevance models. In: ACM SIGIR Conference (2003)
9. Jin, R., Chai, J.Y., Si, L.: Effective automatic image annotation via a coherent language model and active learning. In: ACM Multimedia Conference (2004)
10. Lavrenko, V., Manmatha, R., Jeon, J.: A model for learning the semantics of pictures. In: Neural Information Processing Systems (2003)
11. Li, J., Wang, J.Z.: Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9), 1075–1078 (2003)
12. Liu, X., Jia, Y., Chen, X., Deng, Y., Fu, H.: Image classification using the max-min posterior pseudo-probabilities method. Technical Report BIT-CS-20080001, Beijing Institute of Technology (2008), `http://www.mcislab.org.cn/member/~xiabi/papers/2008_1.PDF`
13. Monay, F., Gatica-Perez, D.: On image auto-annotation with latent space models. In: ACM Multimedia Conference (2003)
14. Mori, Y., Takahashi, H., Oka, R.: Image-to-word transformation based on dividing and vector quantizing images with words. In: Workshop Multimedia Intelligent Storage and Retrieval Management (1999)
15. Perronnin, F.: Univeral and adapted vocabularies for generic visual categorization. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(7), 1243–1256 (2008)
16. Feng, R.M.S., Freitas, D.: Multiple bernoulli relevance models for image and video annotation. In: IEEE Conference on Computer Vision and Pattern Recognition (2004)
17. Wang, X., Zhang, L., Li, X., Ma, W.: Annotating images by mining image search results. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(11), 1919–1932 (2008)
18. Winn, J., Criminisi, A., Minka, T.: Object categorization by learned universal visual dictionary. In: IEEE International Conference on Computer Vision (2005)
19. Zhang, R., Zhang, Z.: Effective image retrieval based on hidden concept discovery in image database. IEEE Transactions on Image Processing 16(2), 562–572 (2007)