Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1
© 2005 - 2013 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SCENE CLASSIFICATION BASED ON THE CONTEXTUAL SEMANTIC INFORMATION OF IMAGE

[1] GUANGYU WEN, [2*] YAN TANG, [3]MENGDIE WU, [4]ZHIXING HUANG
[1,2,3,4] College of Computer and Information Science, Southwest University, Chongqing, China
E-mail: [1]guangywen@gmail.com , [*2]ytang@swu.edu.cn , [3]wmd1124@gmail.com, [4]huangzx@swu.edu.cn

ABSTRACT

Scene classification is an important research direction in the computer vision. However, it is not an easy task. We face many serious difficulties and challenges when classifying the nature scenes. A novel approach is proposed to recognize the nature scenes. Based on the traditional Bag of Visual words (BOV) model, the feature field and space field are combined by introducing the Markov Random Field (MRF) when quantifying the image into a collection of unordered visual words. And then the Latent Dirichlet Allocation (LDA) model is applied to learn the topic distribution of scenes. At last, the Support Vector Machine (SVM) is used to build a classifier in order to categorize a new image. The experimental results on the dataset of 15 nature scenes demonstrate that the introduction of the contextual semantic information on the basis of the traditional method can improve the classification performance.

Keywords: *Scene Classification, Markov Random Field, LDA, Bag of Visual Words*

1. INTRODUCTION

The goal of scene classification is to annotate automatically the scene images with a semantic label (such as coast, living room, etc.) according to the image contents. With the dramatic increase in the number of the scene images, how to classify these massive amounts of images effectively has become a hot spot in the computer vision.

Three strategies can be found in the literatures. The first categorizations are based on the global features. The visual content of the image is depicted by the global features which are extracted from the image (such as shape, color, etc.). More and more scholars fuse a variety of low-level features to make the representation of scenes content more rational since the single low-level features are insufficient to reliable categorization [1][2]. Because these methods are sensitive to the image content, so they are not suitable for complex scenes. The second categorizations are based on the local target regions of scenes. Vailaya etc.[3] categorized the scenes into some simple categories(such as indoor VS outdoor, city VS country, etc.) . They segmented the image into some meaningful target regions (such as sky, sea, etc.), and then recognized the semantic category of the scene according to the contents of the target regions. These approaches recognize the objects first and then in turn identify the category of scene. However, these approaches are very time-consuming and tedious in that they require not only a large scale of training set but also the automatic or manual annotations of massive local regions. Furthermore, the classification accuracy of these approaches is not desirable due to the hand-annotations are arbitrary and ambiguous. Neither the approach based on the global features nor the approach based on the local region can obtain a satisfactory classification precision in view of the semantic "gap" between the low-level features and the high-level semanteme. The third classifications are based on the "Bag of Visual words" (BOV) model. Firstly, the local features are extracted from the scene to depict the semantic conception in order to generate the visual words. Secondly, each image is quantified into a collection of unordered visual words. Lastly, some machine learning algorithm (such as SVM) is applied to recognize the scenes. On the groundwork of the BOV model, Bosch[4] and Feifei Li[5] improved the classification performance by resorting to PLSA[6] and LDA[7] in the text processing literatures respectively. All these approaches took adequately the feature field of images into account. They retained the frequencies of visual words via mapping the local feature to the nearest cluster center (visual word) when quantifying the image into a collection of unordered visual words. However, such a mapping algorithm discarded all the information in spatial

*corresponding author

© 2005 - 2013 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

layout. They considered only the positional relationship among the local features in the feature field, but ignored the contextual semantic information in the spatial field. It was no doubt that some deviations existed in the descriptions of the images. And these deviations affected in turn the classification performance.

A scene classification based on the contextual semantic information of image is proposed to overcome the lack of traditional BOV model. The Markov Random Field (MRF) is introduced into the traditional BOV model. It makes the generation of the visual-words frequency matrix more reasonable by considering not only the Euclidean distance(ED) from the local feature to the visual word but also the contextual semantic information among the local features when mapping the local feature to the visual word. The experimental results show the introduction of the contextual semantic information can enhance the classification accuracy.

2. MARKOV RANDOM FIELD MODEL

Markov Random Field (MRF)[8] is the extension of the Markov Random Chain (RFC) on the 2-dimensional dataset. It is very suitable to model the states of the pixel grid in the image because it can finely depict the inter-action and inter-dependence relationship among the adjacent sites in the lattice.

A random process is said to be a Markov Random Field if the following condition holds. The conditional probability function for the state $Z_{i,j}$ at a site $s=\{(i,j)|1\le i,j\le n\}$ given the states of all other sites on lattice S is equal to the conditional probability for that state given only the states in the neighborhood system of site s. Following the notation of Li's book [9], this can be written as follows:

$$p(Z_{i,j}|Z_{S-\{i,j\}})=p(Z_{i,j}|Z_{N(i,j)}) \quad (1)$$

Where S is a set of sites on a square $n\times n$ lattice, $S-\{i,j\}$ refers the other sites in the lattice S in addition to the site $s_{i,j}$, $N(i,j)$ is called neighborhood system of site *s* which is the set of sites neighboring site *s*. Clearly, the sites in *S* are related to one other via a neighborhood system. Usually, the neighborhood can be 1st order or 2nd order.

According to the Hammersley–Clifford theorem [10], a Markov Random Field can be written as follows if and only if its configurations obey a Gibbs Distribution:

$$p(Z_{i,j})=Z^{-1}\times exp\left(-U(Z_{i,j})\right) \quad (2)$$

Where $Z=\sum exp\left(-U(Z_{i,j})\right)$ is the partition function to normalize; $U(Z)=\sum_{c\in C}V_c(Z)$ is the energy function, c is the clique and $V_c(Z)$ is the clique potential function of c. There are several choices for the clique potential function $V_c(Z)$, and Besag[11] defined the $V_c(Z)$ as follows:

$$V(Z_{i,j},Z_{p,q})=-\beta\delta(Z_{i,j},Z_{p,q}) \quad (3)$$

Where $\delta(Z_{i,j};Z_{p,q})=\begin{cases}1 & if\ Z_{i,j}=Z_{p,q}\\ 0 & else\end{cases}$, β is the parameter which controls the interaction strength in the neighborhood system.

3. OUR APPROACH

"Bag of Visual words" (BOV) model receives a widely attention in recent years because it can reduce the semantic "gap" between the low-level feature and the high-level semanteme via an intermediate latent "theme" representation. Traditional BOV model represents the scene image as a collection of unordered visual words by the following steps: (1) Extract the low-level features (SIFT feature) from the image. (2) Form the codebook by clustering all the SIFT features into M clusters and each cluster is seen as a visual word. (3) Quantify the image into a collection of unordered visual words which are from the large vocabulary of codebook---i.e. the visual word frequency matrix of the image. Traditional BOV model mapped all the SIFT features in the same cluster to the center of that cluster when gathering the statistic of the visual word frequency matrix of image. However, this phase of generating the visual word frequency matrix considered only the location relationship among the features in the feature field but ignored the contextual semantic relationship among the features in the spatial field. Indeed, the contextual semantic information contained in the spatial field can reflect some spatial organization in the image. For example, when "grass" appears in the image, the probability to capture the word which has the same meaning as "grass" in the neighborhood is much higher than the probability to capture the words which has other meaning. Obviously, the semantic distribution in the spatial field of image meets certain semantic information constraints.

It is justified to classify the scenes according to the characters in both feature field and spatial field. To this end, we introduce the Markov Random Field into the traditional BOV model. The image is quantified into a more reasonable visual word frequency matrix by combining both the feature field and the spatial field when mapping the low-

Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1
© 2005 - 2013 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

level feature to the visual word. The specific classification framework is shown in Figure 1. The modules of solid line represent the training phase and the modules of dotted line represent the testing phase.

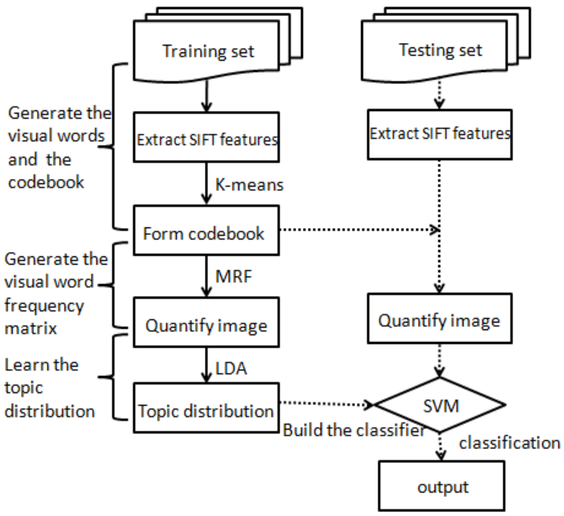


Figure1. Framework Of The Scene Classification Based On The Contextual Semantic Information

3.1 Generate The Visual Words And The Codebook

The "Bag of Visual words" (BOV) model is the extension of the "Bag of Words" (BOW) in the image categorization field. The image set is treated as the corpus, the image as a document and the low-level feature (called visual word) as the word. In this paper, we choose the SIFT feature [12] as the low-level feature of scene. The image is represented as a collection of several 128-D vectors which are extracted from the image. If we regard each SIFT descriptor as a visual word, the vocabulary of codebook will too redundant to cause a serious phenomenon of "different words with the same meaning". To avoid this phenomenon, it usually cluster the SIFT vectors by resorting to some unsupervised clustering algorithm. The SIFT features are clustered into M clusters by performing Kmeans algorithm and each cluster is seen as a visual word. Through this, we reduce the dimension of the large number of SIFT features and form a codebook with a capacity of M visual words.

3.2 Generate The Visual Word Frequency Matrix

In the traditional BOV model, all the SIFT features in the same cluster which are the results of the Kmeans algorithm are mapped to the same visual word. Such a mapping algorithm is so dependent on the positional relationship of the SIFT features in the feature field that the visual words are semantic ambiguous. It will lead to the phenomenon of "the same word with the different meaning" if the SIFT features with different semanteme are mapped to the same visual word according only to their short Euclidean distance in the feature field. Similarly, It will result in the phenomenon of "the different words with the same meaning" if the SIFT features with the same semanteme are mapped to the different visual words according only to their long Euclidean distance in the feature field.

To resolve the foregoing problem, we introduce the Markov Random Filed into the traditional BOV model in order that we can take advantage of the contextual semantic information to model the image. The statistical information based on the contextual semantic constraint is added into the traditional method.

Let $F = \{\cup f_{i,j} | 1 \le i,j \le n\}$ be the SIFT features set which are extracted from the images and distributed in the lattice *S*, where *i* and *j* are the coordinates of the feature. $Z(f_{i,j}) \in \{1,2,\dots,M\}$ is the cluster ID assigned to each SIFT feature after Kmeans algorithm. The interaction and inter-dependence among the sites are confined to the neighborhood system $N(i,j)$. According to the Hammersley–Clifford theorem [10], the conditional probability $p\left(Z_{i,j} \middle| N(i,j)\right)$ for state at the site $s(i,j)$ given the neighborhood system $N(i,j)$ is derived as follows:

$$p\left(Z_{i,j} \middle| N(i,j)\right) p\left(Z_{i,j} \middle| N(i,j)\right) = \frac{\exp\left(\sum_{c \in C} \beta\delta(Z_{i,j}; Z_{p,q})\right)}{\sum_{N(i,j)} \exp\left(\sum_{c \in C} \beta\delta(Z_{i,j}; Z_{p,q})\right)} \tag{4}$$

Where $\delta(Z_{i,j}; Z_{p,q}) = \begin{cases} \frac{1}{N} & if\ Z_{i,j} = Z_{p,q} \\ 0 & else \end{cases}$ which is modified on the basis of the Besag's[11] definition and *N* is the total number of the SIFT features in the neighborhood system of the site $s(i,j)$.

Use the formula (4) to improve the traditional mapping function:

$$ds' = \frac{ds}{p(Z_{i,j}|N(i,j))} \tag{5}$$

Where ds is the Euclidean distance from the local feature to the center of each cluster; And (i,j) are the coordinates of the feature in the image. $N(i,j)$ is the neighborhood system of the site $s(i,j)$. Because the position of the SIFT feature vector is actually a pixel point, so we need to modify the traditional neighborhood system. The definition of the neighborhood system $N(i,j)$ of the site $s_{i,j}$ in

Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1
© 2005 - 2013 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

this paper is $N(i,j) = \cup f_{p,q}$, where $f_{p,q}$ is the local feature and $p \in [i-8, i+8] \&\& q \in [j-8, j+8]$.

In this paper, we utilize the formula (5) to update the distance from the SIFT feature vector to the center of each cluster. It is more reasonable than the traditional approach. Obviously, the probability for the site *s* with the same meaning as "sea" is much higher than it with the other meaning when the most sites in the neighborhood system of *s* are the meaning of "sea". The improved algorithm takes into account not only the position information in the feature field but also the position information in the spatial field among the feature vectors. In the improved mapping algorithm, the site *s* can be mapped to the *k*th visual word when the most of the sites in its neighborhood are mapped to the *k*th visual word even if the Euclidean distance from s to the *k*th visual word is not the nearest. The symbiotic phenomena in the contextual semantic information in the spatial field are added into the traditional mapping algorithm by combining the feature field and spatial feature. The model learn the distribution of each visual word and form the visual word frequency matrix after mapping each SIFT feature to the fittest visual word.

3.3 Learn The Topic Distribution

We resort to the Latent Dirchlet Allocation (LDA)[7] to learn the potential topic distribution of each scene image. LDA is a generative model which is used on the discrete dataset and has been widely used in the field of nature language processing and the image classification [4][5] on account that it can reduce the dimension of the high dimensional dataset on a large scale and in no loss of semantic information. We make use of the LDA to learning the potential topic distribution of each image. Then each scene image is represented as a mixture distribution of several potential topics and each topic is represented as a mixture distribution of several visual words.

Just as the same as the generative process of LDA in text processing field, the generative process of generating an image is just as follows:

The LDA generative mode of an image

1.choose $\boldsymbol{\vartheta} \sim \boldsymbol{Dirichlet\ (\alpha)}$, where $\boldsymbol{\alpha}$ is the prior parameter of the Dirichlet distribution; $\boldsymbol{\vartheta}$ is a $C\times T$ matrix and the row $\boldsymbol{\vartheta_i}$ represents the potential topic distribution of the *i*th image.

2.for the each patch $\boldsymbol{x_i}$ in the image:

a. sample the topic $\boldsymbol{t_k}$ from the polynomial distribution $\boldsymbol{\theta}$ and $\boldsymbol{t_k \sim Multi\ (\vartheta)}$

b. choose a visual word $\boldsymbol{w_n}$ for the probability of $\boldsymbol{p(w_n|t_k,\beta)}$, $\boldsymbol{\beta}$ is a $K\times V$ matrix, the element $\boldsymbol{\beta_{i,j} = p(w_i = 1|t_j = 1)}$ represents the probability of the theme $\boldsymbol{t_j}$ and the visual word $\boldsymbol{w_i}$ appearing at the same time.

3.repeat the above steps, constantly choose a topic and in turn choose a visual word according to the topic until a complete image is generated

The learning process of LDA is actually the inverse process of the LDA generative model. The mixture topic distribution of the images is just as follows:

$$\mathrm{p}(\vartheta, t, w|\alpha, \beta) = p(\vartheta|\alpha) \prod p(t_n|\vartheta) p(w_n|t_n, \beta)$$

We resort to the Gibbs Sampling [13] to estimate the value of the parameters in LDA model. Thus the mixture topic distribution of scene for each scene category is obtained.

3.4 Svm Classifier

We can learn the probability model of the mixture topic distribution for each scene categories by the foregoing approach. SVM is widely used in computer vision [14][15] and applied to category the scenes in this paper. A distinguished SVM classifier can be constructed for each scene class by making use of the mixture topic distribution to train the SVM classifier. The *"one–against-all"* methodology is used to build the multi-class classifier so that to realize the scene classification. Each scene class corresponds to a SVM classifier. In the testing phase, the new image which need to be categorized is extracted the SIFT feature first. Then the new image is represented as a collection of unordered visual words from the large vocabulary of codebook which is learned in the training phase. The mixture topic distribution of the new image is learned by LDA. The Gibbs Sampling is applied to estimate the parameters of LDA in the reasoning phase of LDA so that we can get the mixture topic distribution of the new image. The difference is that the condition probability of the visual word given the topic $p(w|t)$ is fixed. The output confidence of the new image for each scene SVM classifier is calculated separately. The new image is labeled with the category of the SVM classifier which has the highest confidence.

3.5 Scene Classification Based On The Contextual Semantic Information Of Image

To sum up, the specific algorithm of scene classification based on the contextual semantic information of image is elaborated as below:

Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1
© 2005 - 2013 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The Algorithm of Scene Classification Based on the Contextual Semantic Information of Image

1. extract the SIFT features from the images of all categories $F = \{\bigcup f_{i,j} | 1 \leq i, j$
2. cluster all the SIFT features into M clusters by performing Kmeans algorithm in order to form the codebook
3. calculate the Euclidean distance from the SIFT feature to the center of cluster for each SIFT feature
4. update the distance from the SIFT feature to the center of cluster:

$$ds' = \frac{ds}{p(Z_{i,j}|N(i,j))}$$

5. map the SIFT feature to the corresponding visual word according to the improved distance and generate the visual word frequency matrix of scene image.
6. let the visual word frequency matrix be the input of the LDA model in order to learn the potential topic distribution of each scene category
7. construct the SVM classifier for each scene categories with the *"one-against-all"* methodology
8. apply the SVM classifier to identify the new scene image

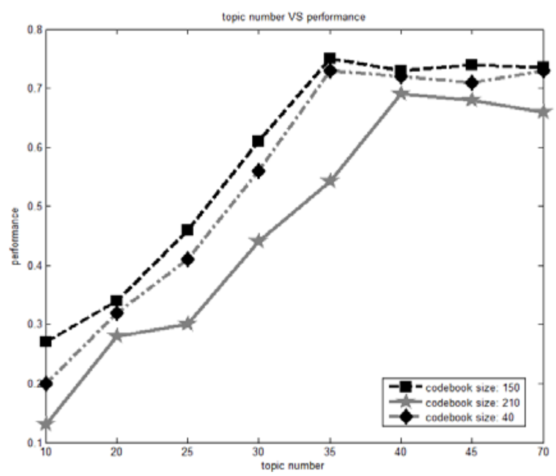
4. Dataset and experimental setup

The experimental dataset1 is the *state-of-art* 15 nature scenes which are also used in [16-18]. The dataset contains 15 categories of natural scenes and the pixel size of each scene is about *256×256*. We turn all the color images into grayscale images even though the color images are available.

All the categories of scenes are split randomly into two separate sets of images, 100 for training and the rest for testing. The *"one-against-all"* methodology is used to build the SVM classifier. For each image in testing set, the output confidence of every scene classifier is calculated separately and the label of SVM classifier with the highest confidence is the category of the new image.

4.1influencing Factors Of Classification Accuracy

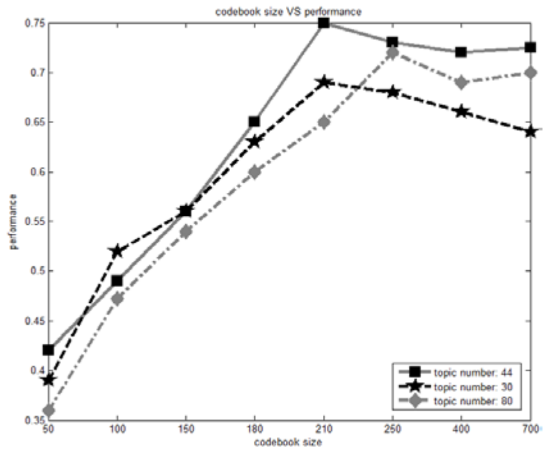
The experiment is repeated with different parameter setting in order to make the influencing factors of classification accuracy clear. The influence of the topic number and codebook capacity on the average classification performance is displayed in Figure 2. The Figure 2(a) shows how a change in the number of topic will influence the average classification accuracy when the codebook capacity is fixed while the Figure 2(b) illustrates that how a change in the capacity of the codebook will influence the mean classification accuracy when the number of topics is fixed. From Figure 2(a), we can find that the increase of the average classification performance is very significant at first with the increase of the number of the topics. And the average classification performance reaches the highest when the number of topic is 40. As the number of topic further increase, the average classification performance decline slowly. Similarly, we can obverse from Figure 2(b) that the accuracy gradually increases with the increasing of the capacity of the codebook at the beginning, and reach the highest when the capacity of codebook is 210, but if the capacity of codebook increases further, the accuracy will be descend lightly. The reason for such a trend maybe a too small capacity or the topic number will lose too much semantic information, while a too large capacity or the topic number will cause too much semantic redundancy to affect classification performance.



(A)

1 http://www-cvr.ai.uiuc.edu/ponce_grp/data/

Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1
© 2005 - 2013 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

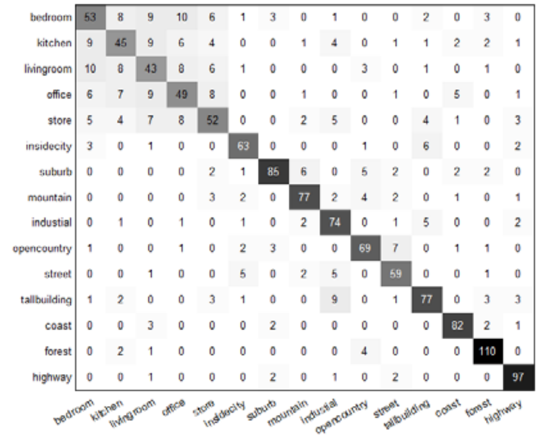


(B)

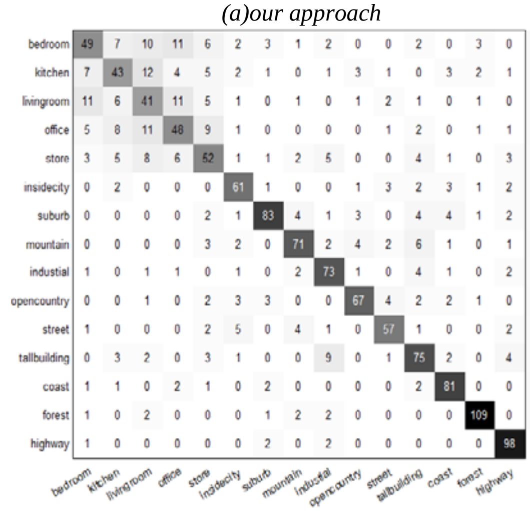
Figure2. The Influence Of The Topic Number And The Codebook Size On Classification Accuracy

4.2 Confusion Matrix For The 15 Nature Scene

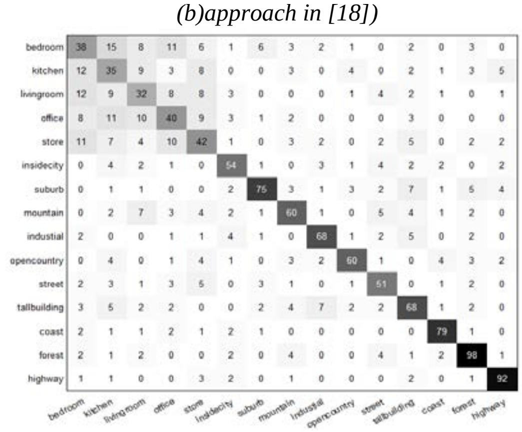
Figure 3 shows the confusion matrix of 15 nature scenes which can illustrate the performance of each category in three approaches separately. Figure 3(a) is the confusion matrix of our approach while the Figure 3(b) and the Figure 3(c) are the confusion matrix of the approach in [18] and [5] separately. The *x-axis* and *y-axis* both represent a scene class and the order is consistent. The entry in the *i*th row and *j*th column is the number of the scenes from the class *i* that are misidentified as class *j*. The depth of the color in the confusion matrix can reflect the number of images which is classified to each class. The darker color the more images are categories to the class. Conversely, the lighter color the less images are identified to the class. Clearly, an ideal confusion matrix should be a diagonal matrix where the diagonal color should be black while the color of the other elements should be white. For a given scene with the fixed number, the diagonal entry with the bigger value or the darker color indicates that the misidentified scenes are less than the diagonal element with the smaller value or the lighter color. It can be easily found from the Figure 3 that the color or the value of the diagonal elements in Fig3(a) are much blacker or bigger than that in Fig3(b) and Fig3(c). It illustrates that our approach can obtain a higher accuracy than the other approaches. Taking a close look at the confusion matrix, we can find that ,for all the three confusion matrix, the accuracy of the five indoor categories (bedroom, living room, kitchen, office and store) are low, indicating that these five categories are confused to each other. Therefore, a further improve algorithm is needed to enhance the classification accuracy of the indoor scenes.



(a)our approach



(b)approach in [18])



(c)approach in [5]

Figure3. The Confusion Matrix Of The 15 Nature Scenes For Three Approaches

4.3 Classification Accuracy

The highest average classification accuracy (75.04%) is obtained when the size of the codebook is 210 and the topic number is 40 after several repeated experiments. The Table 1 exhibits the average classification performance of the traditional

Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1
© 2005 - 2013 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BOV model in [5] and [18] as well as our approach separately.

Table 1. The Average Classification Accuracy

approach	*The classification performance for varied categories (%)*			
	15	13	8	4
Our approach	***75.04***	***77.64***	***81.27***	***87.87***
Approach in[18]	73.01	75.35	81.23	87.79
Approach in[5]	64.50	69.41	75.70	83.82

The specific accuracy for each scene is shown in Figure 4. It can be found that our method improves the precision of classification for the most categories. Be consistent with the study of [5] and [18], some outdoor scenes (especially forest and coast) in all the three methods can achieve a satisfactory classification results. Maybe the reason is that the outdoor scenes are usually high texture. Although we enhance slightly the classification accuracy of the five indoor categories, but the results are far from satisfactory. Perhaps, the reason is that the indoor scenes are usually single and the too much similar semantic information reduces the distinguish ability from each other and in turn affect the classification performance. For instance, the "cabinet" patch may appear in all the five scenes. Therefore, the distinguished ability from each other is abated and in turn affects the classification performance ultimately.

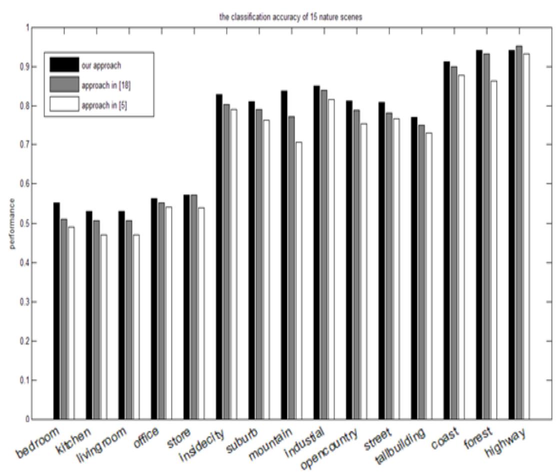


Figure4. The Classification Accuracy For Each Category In Our Approach/[18]/[5]

5. Discussion

In this paper, a novel approach is presented to classify the scenes based on the contextual semantic information. We reduce the "gap" between the low-level feature and high-level semanteme to a certain extent by introducing the Markov Random Field and improve the classification performance. Our approach outperforms the traditional BOV model. And the experimental results of our approach are more precious than the traditional BOV model on the dataset of 15 nature scenes proving the feasibility and effectiveness of our approach. However, the discouraging classification accuracy of five indoor classes is still far from satisfactory. How to further improve the classification accuracy of indoor scene is the next work.

ACKNOWLEDGEMENTS
This project is supported by Cunhui Project "Research about text detection and recognition in images and video frames Z2011149"

REFRENCES:

[1] Benmokhtar, R., Huet, B., and Berrani, S. Low-level Feature Fusion Models for Soccer Scene Classification [C]// In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME).2008 June.1329-1332
[2]Anne H.H.Ngu, Quan Z.Sheng,Du Q.huynh, Ron Lei, Combining multi-visual features for efficient indexing in a large image database, The VLDB Journal-The International Journal on Very Large Data Bases,2001,April,9(4);279-293
[3]Vailaya, A., Figueiredo, M. A. T., Jain, A. K., and Zhang, H.J. Image classification for content-based indexing. IEEE Trans. Image Process.2001,10(1) 117--130.
[4]Bosch A, Zisserman A,Muoz X. Scene classification using a hybrid generative /discriminative approach [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, Apr.2008，30(4): 712-727
[5]Li Fei-Fei,P Persona.A Bayesian hierarchical model for learning natural scene categories [A]//Proc. Of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR05) [C].USA,IEEE Computer Society, 2005(2): 524- 531
[6]Hofmann T. Unsupervised learning by probabilistic latent semantic analysis[J]. Journal of machine Learn- ing, 2001,41(2):177-196.
[7]Blei D, Ng A,Jordan M. Latent Dirchlet Allocation [J].Journal of Machine Learning Research, 2003,2(3): 993-1022
[8]Kindermann, Ross; Snell, J. Laurie. Markov Random Fields and Their Applications [M]. Amer. Math. Soc., Providence, R.I.1980

© 2005 - 2013 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[9]Stan Z. Li. Markov Random Field Modeling in Image Analysis [J], Computer Science Workbench, Springer,2001.

[10]Geman S, Geman D.Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984 (6):721-741

[11]Besag J. On the Statistical Analysis of Dirty Pictures [J]. Journal of Royal Statistical Society, 1986, B48(3): 259-302

[12] David GLowe. Distinctive image features from scale- invariant key points [J].International Journal of Computer Vision. 2003, 60(2):91-110

[13]Porteous Ian, Arthur Asuncion, David Newman. Fast collapsed Gibbs sampling for latent Dirichlet Allocation [C]//In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008:569-577

[14]Phd Student, RT-LARIT, FSK, Morocco. Finger-Knuckle-print recognition based on local and global feature sets. Journal of Theoretical and Applied Information Technology. December 2012. 46(1): 54-61

[15]K.NARASIMHAN, NEHA. V. C, K. VJAYAREKHA. A review of automated diabetic retinopathy diagnosis from fundus image. Journal of Theoretical and Applied Information Technology. May 2012,39(2): 225-239

[16]S Lazebnik,C Schmid, J Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories.CVPR, 2005: 2169 - 2178

[17]Nikhil Rasiwasia, Nikhil Rasiwasia .Scene Classification with Low-dimensional Semantic Spaces and Weak Supervision[A]// Proc. Of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR ’ 08)[C],June 2008: 1-6

[18] J. Liu and M. Shah, Scene Modeling Using Co-Clustering, Proc. Conf. Computer Vision, 2007