Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2015, Article ID 109718, 14 pages
http://dx.doi.org/10.1155/2015/109718

Research Article

A New Scene Classification Method Based on Local Gabor Features

Baoyu Dong[1,2] **and Guang Ren**[2]

[1] *College of Electric Information, Dalian Jiaotong University, Dalian 116028, China*
[2] *Marine Engineering College, Dalian Maritime University, Dalian 116026, China*

Correspondence should be addressed to Baoyu Dong; signaldong@163.com

Received 10 November 2014; Revised 9 March 2015; Accepted 6 April 2015

Academic Editor: Lucian Busoniu

Copyright © 2015 B. Dong and G. Ren. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A new scene classification method is proposed based on the combination of local Gabor features with a spatial pyramid matching model. First, new local Gabor feature descriptors are extracted from dense sampling patches of scene images. These local feature descriptors are embedded into a bag-of-visual-words (BOVW) model, which is combined with a spatial pyramid matching framework. The new local Gabor feature descriptors have sufficient discrimination abilities for dense regions of scene images. Then the efficient feature vectors of scene images can be obtained by K-means clustering method and visual word statistics. Second, in order to decrease classification time and improve accuracy, an improved kernel principal component analysis (KPCA) method is applied to reduce the dimensionality of pyramid histogram of visual words (PHOW). The principal components with the bigger interclass separability are retained in feature vectors, which are used for scene classification by the linear support vector machine (SVM) method. The proposed method is evaluated on three commonly used scene datasets. Experimental results demonstrate the effectiveness of the method.

1. Introduction

Scene classification is an appealing and challenging problem in image processing and machine vision. The goal of scene classification is to automatically classify scene images into specific scene categories such as mountain, street, forest, and inside city. Scene classification methods have many applications, such as video retrieval, content-based image retrieval, UAV autonomous landing, and intelligent vehicle navigation [1]. Moreover, scene classification can provide an important cue for object recognition and detection, action recognition, and other computer vision tasks.

Scene classification methods can be divided into two main categories. First, the early methods mainly use low-level global features (e.g., texture and color) which are extracted from a whole image [2, 3]. These methods often exhibit poor classification performance, because they lack an intermediate image description that is extremely valuable in determining the scene category. Second, the methods make use of semantic models [4]. They describe the contents of scene images by the semantic intermediate representation, which can be mainly divided into the local semantic concepts based intermediate representation methods and the global semantic concepts based intermediate representation methods.

The local semantic concepts based intermediate representation methods make use of the features extracted from local regions in scene images [5, 6]. They generally represent the scene image by a collection of local descriptors using segmentation, dense sampling patches, or interest point detectors. These methods are widely used due to their effectiveness, especially the bag-of-visual-words (BOVW) model [7, 8]. The BOVW model extracts local feature descriptors of scene images and obtains visual words by clustering and then uses the histograms of visual words to represent images. The BOVW model has obtained good performance, but this technique also has some limitations. The BOVW model uses the orderless collection of local descriptors to represent scene images [9], and therefore any spatial relationships of scene images are lost. The loss of spatial position information affects the accuracy of scene classification [10]. The weakness of the BOVW model can be mitigated by a spatial pyramid matching framework [11]. In the pyramid matching framework, a scene image is partitioned into increasingly finer grids. The histogram of visual words inside each subregion is computed. The pyramid matching framework has obtained encouraging performance. Nowadays, many of the best scene categorization methods are based on this scheme.

The global semantic concepts based intermediate representation methods take the scene image as a whole for obtaining global description features. The "Gist" model is the most prominent one of these methods. It has exhibited good performance in many applications [12, 13]. In this model, scene images are convolved by the multiscale and multiorientation Gabor filters. Then the filtering results are divided into a $4 * 4$ grid and the means of all subregions are computed and assembled for yielding feature vectors [14]. Lastly, the "Gist" features are used for scene classification. The "Gist" model is obtained from the sparse grid of the scene image. Thus, the "Gist" feature is coarse-grained, and some detailed information of the scene image is lost. When scene images are complex, the classification performance of the "Gist" model is not very good. For example, when some categories of indoor environments are included in scene datasets, the classification accuracy of the "Gist" model drops dramatically.

In this study, we will present a new method for scene classification using local Gabor features. The proposed method not only solves the coarse-grained problem of the "Gist" feature but also utilizes the spatial information of the pyramid matching model. In addition, the proposed method extracts principal components of feature vectors of scene images by the improved KPCA algorithm, which can retain more category information. Last, the linear "1-a-r" SVMs are used for scene classification. For evaluating the performance of the proposed method, three scene datasets are used for classification testing. We also investigate the impacts of different parameters on the performance of the proposed classification method. The proposed method is also compared with several well-known methods.

This paper is arranged as follows: In Section 2, our method of scene classification is described, and the implementation steps are presented. In Section 3, we evaluate the proposed method on three different datasets and present experimental results. In Section 4, the conclusions are given.

2. The Proposed Scene Classification Method

The framework of the proposed method is illustrated in Figure 1. First, scene images are convolved with a 2D Gabor filter bank, and then the image patches of $15 * 15$ pixels are obtained from the filter responses by dense sampling. The local Gabor feature of each sample point is obtained by computing the Gaussian-weighted mean in the corresponding neighborhood of each filter channel and assembling these means in a vector. Accordingly, local Gabor feature descriptors of dense sampling patches of all scene images can be extracted, and then visual words can be obtained by the K-means clustering algorithm. For exploiting spatial position information, the pyramid histogram of visual words (PHOW) based on a spatial pyramid model is used in this scheme. Owing to the relatively high dimension of PHOW, the computational costs of training and testing of SVM classifiers are high. In order to solve this problem and improve classification accuracy, an improved KPCA method is used for extracting appropriate principal components. The feature vectors obtained by the improved KPCA method are used for scene classification by linear SVMs.

2.1. Local Gabor Feature Extraction. Gabor filters are particularly appropriate for obtaining the texture representation of scene images [15]. In this paper, we extract local Gabor features of images for scene classification. Figure 2 illustrates the procedure of feature extraction. Given a scene image, we firstly convolve it with 2D Gabor filters. The 2D Gabor filters [16] are defined as

$$\psi_k(z) = \frac{k_\nu^2}{\sigma^2}\exp\left(\frac{-k_\nu^2 z^T z}{2\sigma^2}\right)\left(\exp\left(ik^T z\right)-\exp\left(-\frac{\sigma^2}{2}\right)\right), \qquad (1)$$

where $z = (y,x)^T$, $k = k_\nu \exp(i\phi) = (k_\nu\cos(\phi), k_\nu\sin(\phi))^T$, $k_\nu = k_{\max}/f^\nu$, $\phi = \mu\cdot\pi/8$, $f = \sqrt{2}$, and $\sigma = \pi$. In this research, we adopt the Gabor filter bank with eight different orientations ($\mu = \{0,1,\ldots,7\}$) and five different scales ($\nu = \{0,1,\ldots,4\}$). The magnitude responses are used for feature extraction. In order to obtain the fine-grained Gabor feature, we perform dense sampling. We utilize 8 pixels as the sampling interval of the dense regular grid. The $15 * 15$ pixel neighborhood of each sample point is used for calculating the local feature descriptor. For each sample point, the Gaussian-weighted mean of the corresponding neighborhood of every channel is computed, respectively. The mean is treated as the feature value of the corresponding filter channel. Then the local Gabor feature descriptor can be obtained by the concatenation of feature values of all channels. The dimension of the local Gabor feature descriptor is 40 ($5 * 8$). By dense sampling, Gabor feature descriptors of 961 sample points can be extracted from a $256 * 256$ scene image.

We use a Gaussian function for weighting calculation in the neighborhood of each sample point. The Gaussian-weighted function is

$$W_{i,j} = e^{-(i^2+j^2)/\gamma}, \qquad (2)$$

where (i, j) denotes the pixel position in the $15 * 15$ neighborhood. The sample point corresponds to $(0,0)$. The pixel in the upper left corner of the neighborhood corresponds to $(-7, -7)$. The pixel in the lower right corner of the neighborhood corresponds to $(7,7)$. γ is the Gaussian width. We let γ be 100 in this study.

The local Gabor feature descriptors are fine-grained Gabor features which have sufficient discrimination abilities for dense sampling patches of scene images. Then we represent scene images using the bag-of-visual-words model. First, we quantize these Gabor feature descriptors into discrete codewords by the K-means clustering algorithm. Each cluster center corresponds to a visual word. Scene images can be represented as histograms of visual words [17] after the Gabor feature descriptors are mapped into visual words.

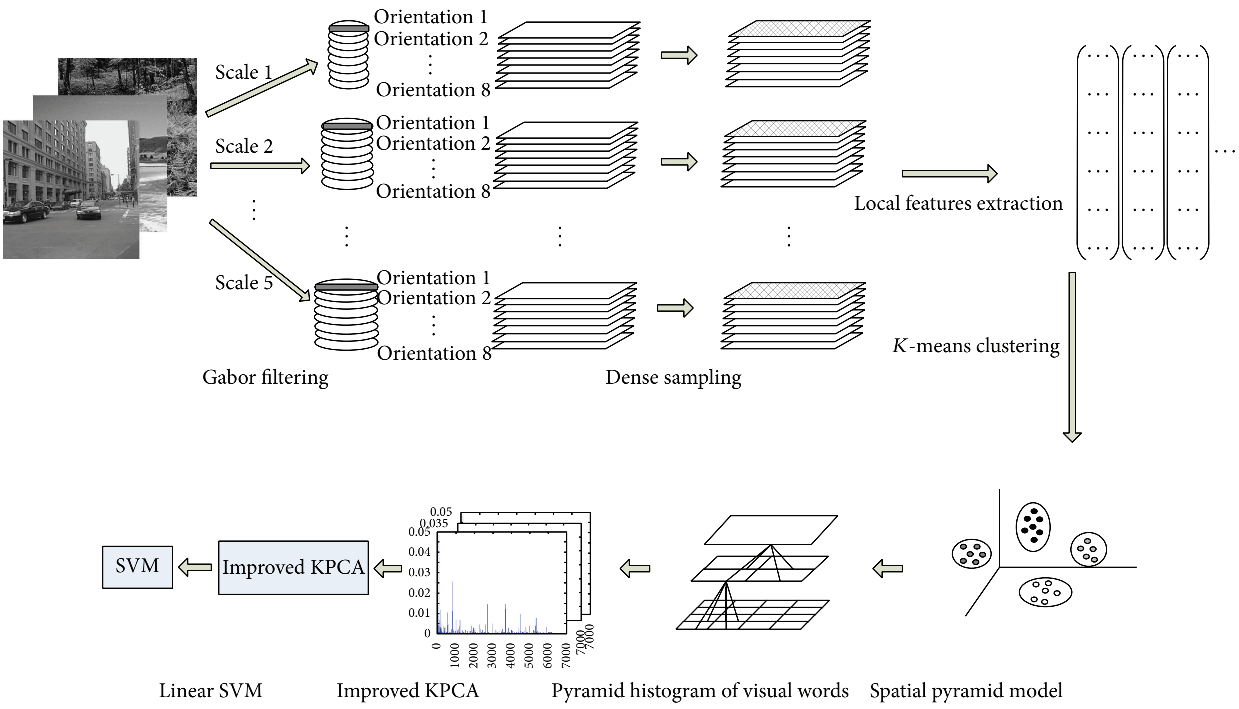


FIGURE 1: Framework of the proposed scene classification method.

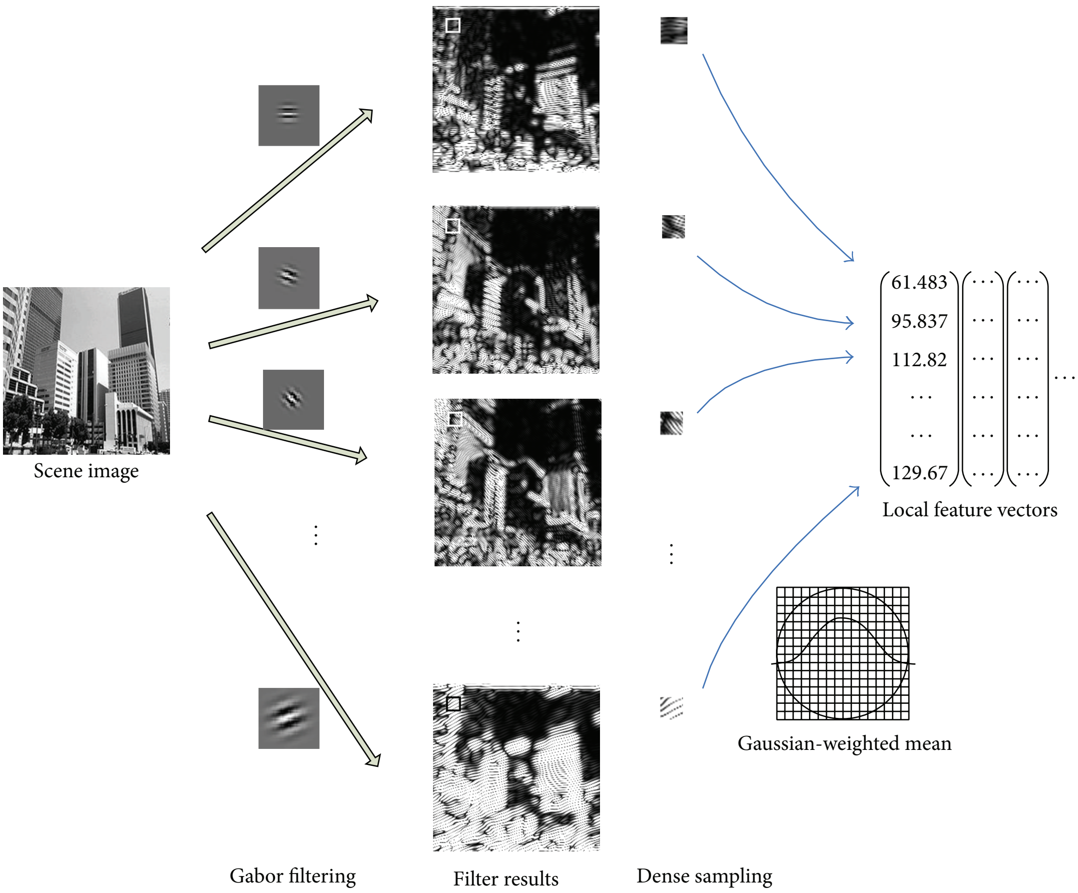
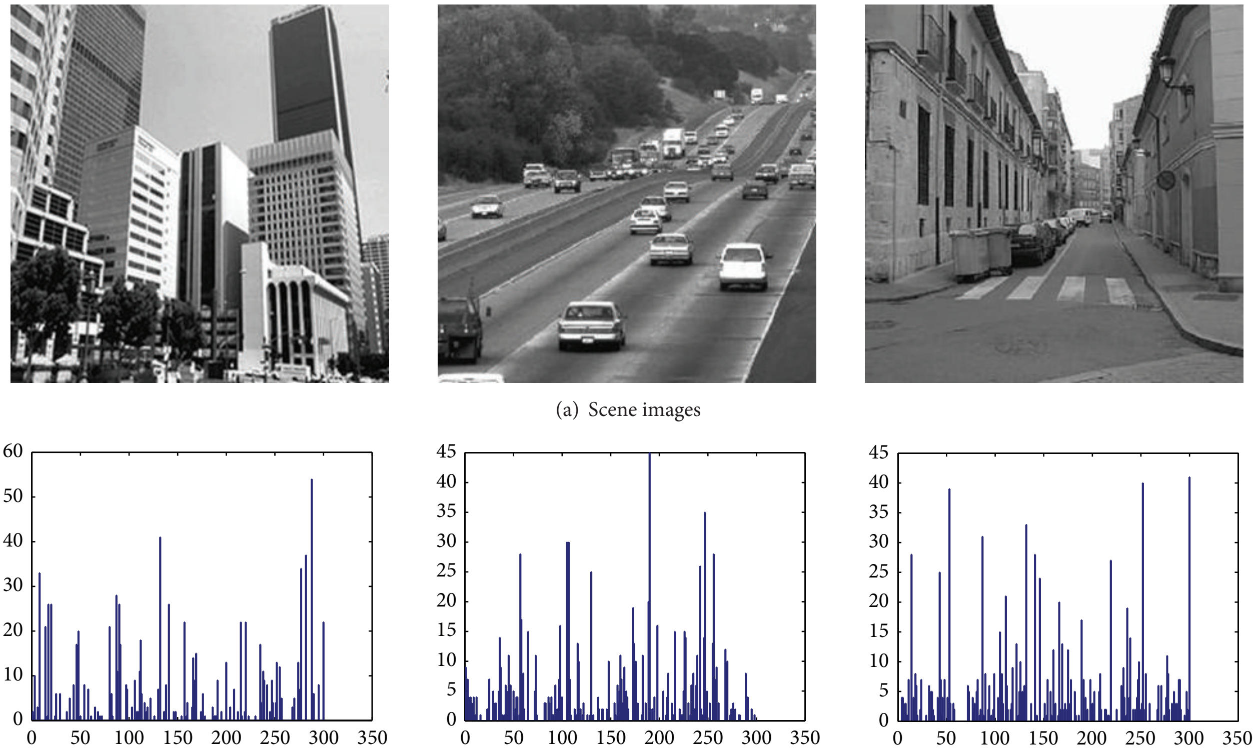


FIGURE 2: Illustration of local Gabor feature extraction.



(a) Scene images

(b) Histograms of visual words

Figure 3: Histogram representation of scene images.

Figure 3(a) illustrates three scene images. Figure 3(b) shows the histograms of visual words based on local Gabor feature descriptors. In this experiment, the vocabulary size is 300. It can be seen that local Gabor features can yield the effective histogram representation of scene images.

We evaluate the local Gabor feature descriptors for scene classification on a 15-category scene dataset [18] and compare them with scale invariant feature transform (SIFT) descriptors [6]. The SIFT descriptors are extracted by dense sampling on a regular grid, which is the same as the grid used by local Gabor features. We use "1-a-r" RBF-SVMs for scene classification and randomly select 200 images of each category as the experiment images. Half of them are used as training samples and the others are used for testing. The codebook size is set to be 300. The comparison results of classification accuracy of all scene categories are shown in Figure 4. We can see that the local Gabor feature descriptor obtains good classification performance. In the same experiment conditions, the classification accuracy of the local Gabor feature is higher than the SIFT descriptor on most of scene categories.

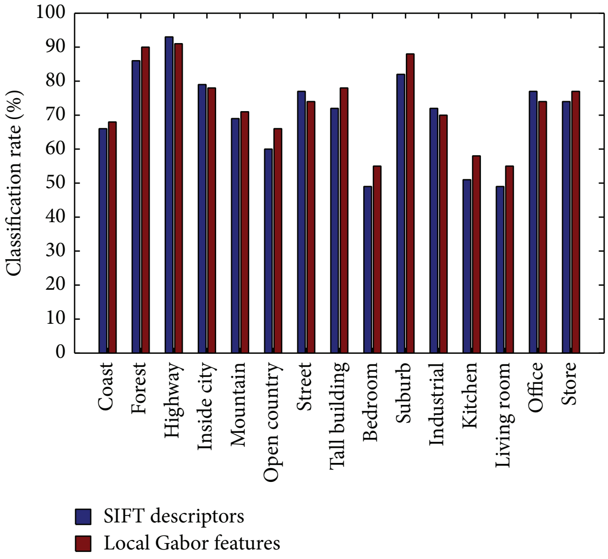


Figure 4: Comparison of classification accuracy using different descriptors.

2.2. Pyramid Histogram of Visual Words (PHOW). The bag-of-visual-words model is limited due to the loss of spatial position information. Thus, we construct a spatial pyramid and compute the pyramid histogram of visual words. The pyramid histogram of visual words is suitable for scene classification because it contains position information of scene images [19]. In order to construct a spatial pyramid, a scene image is partitioned into increasingly finer grids by the quadtree decomposition. A sequence of grids at levels $0, 1, 2, \ldots, L$ are obtained. Then the histogram of visual words inside each subregion is computed, respectively. The PHOW can be obtained by concatenating histograms of visual words of all subregions at different levels.

Figure 5 shows the pyramid histogram of visual words (PHOW) of a scene image. The number of levels of the spatial pyramid is three. For three different levels, the number of visual words of each subregion is counted and shown, respectively. The size of the vocabulary is 300, and therefore the dimensionality of the PHOW is 300 × 21 = 6300.

Using pyramid histograms of visual words as feature vectors for scene classification, a spatial pyramid matching kernel (PMK) is adopted as follows:

$$K(X,Y)=\sum_{m=1}^{M}k(X_m,Y_m), \tag{3}$$

where X and Y represent two scene images and m is the visual word number. $k(X_m,Y_m)$ is defined as

$$\begin{aligned}k(X_m,Y_m)&=I_m^L+\sum_{l=0}^{L-1}\frac{1}{2^{L-l}}\left(I_m^l-I_m^{l+1}\right)\\&=\frac{1}{2^L}I_m^0+\sum_{l=1}^{L}\frac{1}{2^{L-l+1}}I_m^l,\end{aligned} \tag{4}$$

where L is the number of levels and l is the current level. Each level is weighted using $1/2^{L-l}$ for the purpose that matched points from the finer resolution are weighted more highly than those at the coarser resolution. I_m^l is the abbreviation of a histogram intersection function, which is defined as

$$I\left(H_{X_m}^l,H_{Y_m}^l\right)=\sum_{i=1}^{4^l}\min\left(H_{X_m}^l(i),H_{Y_m}^l(i)\right), \tag{5}$$

where $H_{X_m}^l(i)$ denotes the count of the mth visual word in the ith subregion of image X at level l. $H_{Y_m}^l(i)$ denotes the count of the mth visual word in the ith subregion of image Y at level l.

2.3. Improved Kernel Principal Component Analysis.

Pyramid histograms of visual words of scene images are assumed to be x_i $(i=1,2,\ldots,N)$, $x_i\in R^d$. First, KPCA is to map each original input vector x_i into the higher-dimensional feature space H and then compute the covariance matrix:

$$C=\frac{1}{N}\sum_{i=1}^{N}\phi(x_i)\phi(x_i)^T, \tag{6}$$

where here $\phi(x_i)$ is the nonlinear mapping of the input variables x_i. Then we solve the following eigenvalue problem [20]:

$$\lambda V=CV. \tag{7}$$

All solutions V with $\lambda\neq 0$ must lie in the span of $[\phi(x_1),\phi(x_2),\ldots,\phi(x_N)]$ [21], and $V=\sum_{j=1}^{N}\alpha_j\phi(x_j)$. Thus, $\lambda V=CV$ is equivalent to

$$n\lambda\alpha=K\alpha, \tag{8}$$

where K is a $N\times N$ kernel matrix defined by $k_{ij}=K(x_i,x_j)=(\phi(x_i)\cdot\phi(x_j))$. By utilizing the kernel function, nonlinear mapping and inner products computing in the feature space can be avoided [22]. The principal component h_k can be extracted by projecting $\phi(x)$ onto eigenvector V_k as follows [23]:

$$\begin{aligned}h_k(x)=(V_k\cdot\phi(x))&=\sum_{j=1}^{N}\alpha_j^k\left(\phi\left(x_j\right)\cdot\phi(x)\right)\\&=\sum_{j=1}^{N}\alpha_j^kK\left(x_j,x\right).\end{aligned} \tag{9}$$

Let $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_N$ denote the nonzero eigenvalues of the kernel matrix K. By using only the first several eigenvectors sorted in descending order of the eigenvalues, the number of principal components can be reduced [24]. The choice of the number of principal components is as follows:

$$\left(\frac{\sum_{j=1}^{n}\lambda_j}{\sum_{i=1}^{N}\lambda_i}\right)>E, \tag{10}$$

where E is the predefined threshold of the KPCA method.

For simplicity, we have assumed that the observation data are centered, and this could be done by substituting the kernel matrix K with $\widetilde{K}=K-L*K-K*L+L*K*L$, where L is a square matrix. Its elements are all $1/N$.

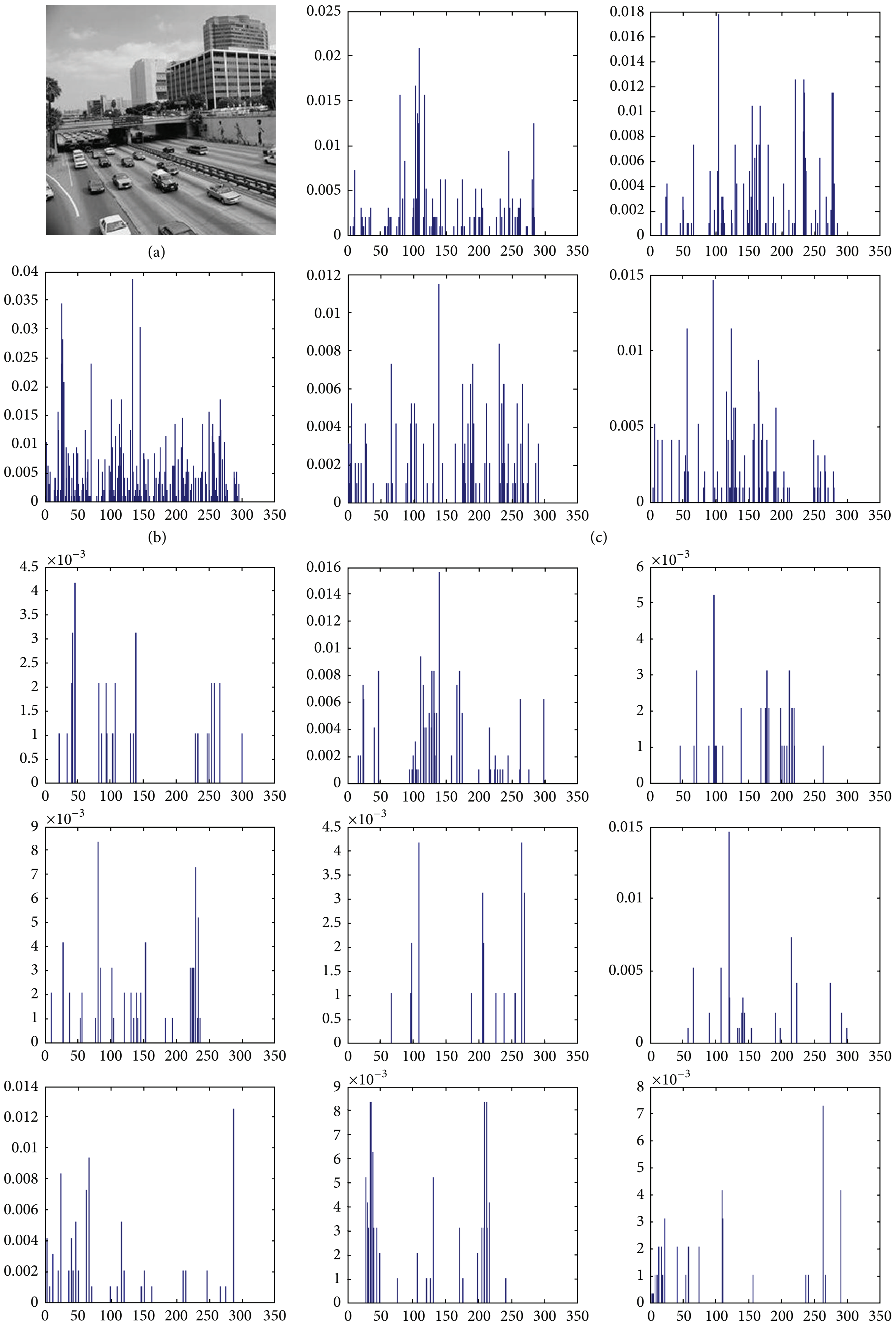
KPCA can retain information as much as possible when feature vectors are simplified. For pattern classification, the most important is not the total amount of retained information but the category information. In view of this, we further extract appropriate principal components by evaluating category information of feature vectors.

In this research, we use the interclass separability for evaluating category information. The separability of the kth dimension component of feature vectors between class i and class j is defined as follows:

$$\delta_{kij}=\frac{d_{kij}}{\sigma_{ki}+\sigma_{kj}}, \tag{11}$$

where d_{kij} is the distance between the center of the kth dimension component of feature vectors of class i and the center of the kth dimension component of feature vectors of class j. Consider $d_{kij}=\|c_{ki}-c_{kj}\|$, where c_{ki} is the center of the kth dimension component of feature vectors of class i. Consider $c_{ki}=(1/N_i)\sum_{l=1}^{N_i}x_{kil}$, where N_i is the number of samples of class i, and x_{kil} represents the kth dimension component of the lth sample of class i. σ_{ki} represents the standard deviation of the kth dimension component of class i. It is formulated as $\sigma_{ki}=\sqrt{(1/(N_i-1))\sum_{l=1}^{N_i}(x_{kil}-c_{ki})^2}$.

The bigger δ_{kij} is, the better the separability of the kth dimension component between class i and class j is. When δ_{kij} is smaller than 1, there is an overlap between the kth dimension component of class i and that of class j.



(a)

(b)

(c)

FIGURE 5: Continued.

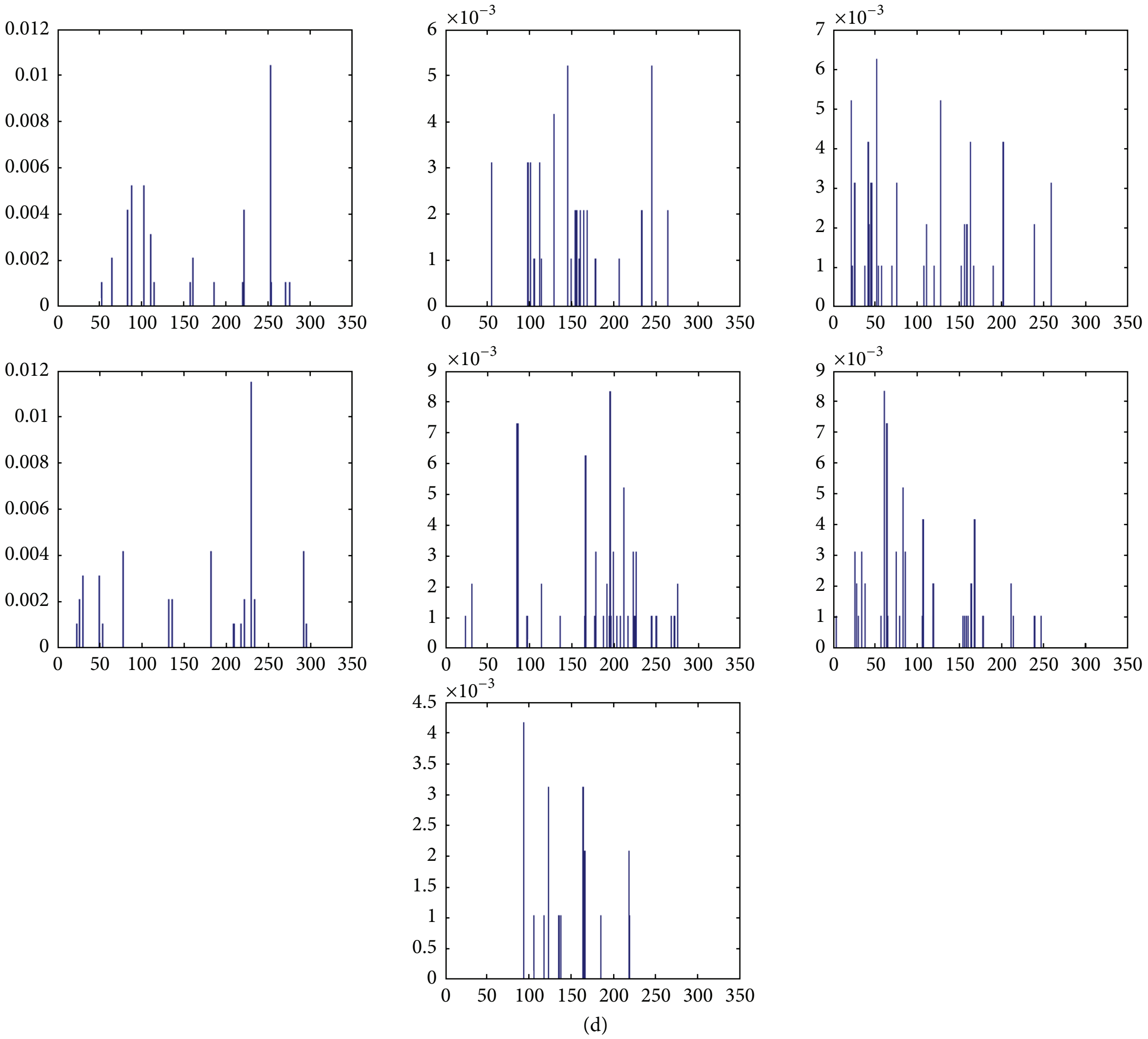


Figure 5: Pyramid histogram of visual words. (a) Scene image. (b) Histogram of level 0. (c) Histogram of level 1. (d) Histogram of level 2.

We define the interclass separability of the kth dimension component of feature vector as follows:

$$J_k = \sum_{i=1}^{C} \sum_{j=i+1}^{C} \delta_{kij}. \tag{12}$$

Let J_k represent the category information of the kth dimension component. The bigger J_k is, the more suitable for classification the kth dimension component is. Then J_k is sorted in descending order, and the components corresponding to the first p separability are retained.

The choice of the number of appropriate principal components for scene classification is as follows:

$$\left(\frac{\sum_{k=1}^{p} J_k}{\sum_{k=1}^{n} J_k} \right) > T, \tag{13}$$

where T is the predefined threshold.

After appropriate principal components are extracted, linear “1-a-r” SVMs [25] are used for scene image classification. The linear SVMs have simple decision function and fast classification speed. These advantages are more prominent for multiclass classification problems.

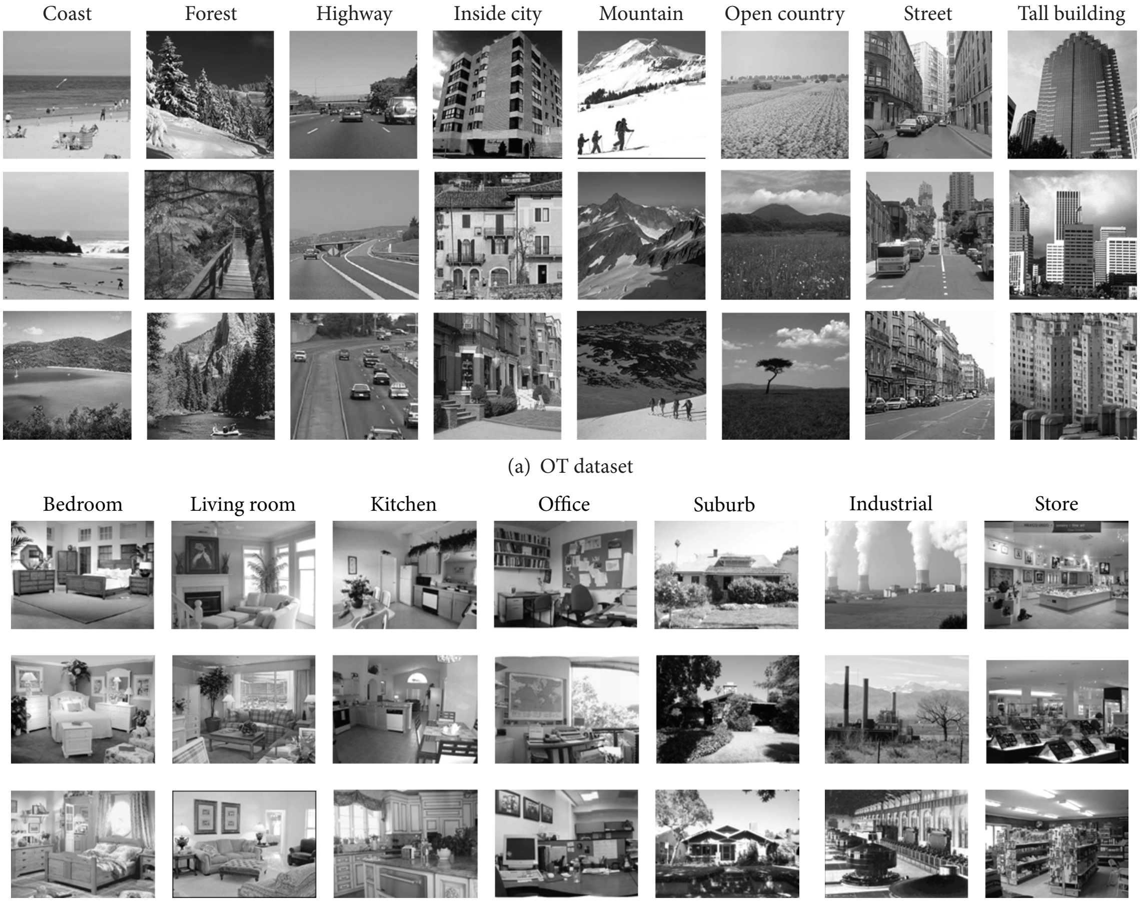
3. Experiments and Results

The proposed method is evaluated on three datasets.

OT dataset [9, 14]: it contains 2688 images from 8 scene categories, which are coast (360 samples), forest (328 samples), mountain (374 samples), open country (410 samples), highway (260 samples), inside city (308 samples), tall buildings (356 samples), and streets (292 samples). The size of each image is 256 × 256.

FP dataset [4, 16]: it contains 3859 images from 13 scene categories. FP dataset is an extension of OT dataset by adding 5 new categories, which are bedroom (216 samples), kitchen (210 samples), living room (289 samples), office (215 samples), and suburb (241 samples). The image size is approximately 300 × 250.

LS dataset [1, 11]: it contains 4485 images from 15 scene categories. LS dataset is an extension of FP dataset by adding



(a) OT dataset

(b) FP dataset

(c) LS dataset

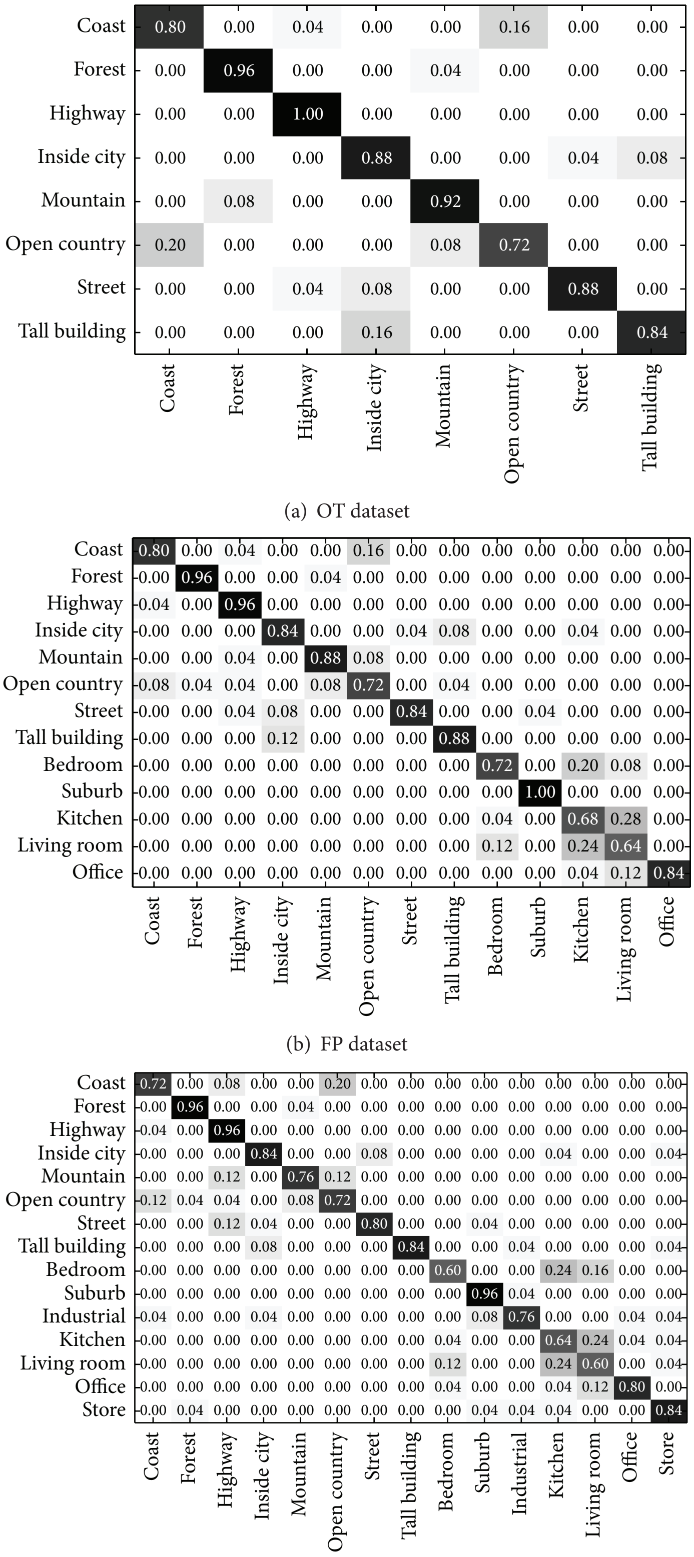
Figure 6: Example images from three datasets.

2 new categories, that is, industrial (311 samples) and store (315 samples). Figure 6 depicts some example images from three datasets. These scene datasets are publicly available at http://www-cvr.ai.uiuc.edu/ponce_grp/data/.

We randomly select 125 images of each category as the experiment images. The fivefold cross-validation is performed for achieving the accurate estimation of classification performance. First, scene images are filtered by the Gabor filter bank of 5 scales and 8 directions, and local Gabor feature descriptors are extracted. Then based on the spatial pyramid matching model, pyramid histograms of visual words are obtained. The vocabulary size is 300, and the number of levels of the spatial pyramid is three. The improved KPCA method with spatial pyramid matching kernel (PMK) is used for dimensionality reduction. The threshold E is set to be 95% and the threshold T is set to be 90%. Last, linear "1-a-r" SVMs are adopted for scene classification. The penalty factor C of the "1-a-r" SVMs is set to be 10.

Figure 7 shows the confusion matrixes of the proposed method for three different scene datasets. In the confusion matrix, average classification rates for individual categories are listed along the diagonal. The entry in the ith row and jth column is the percentage of images from category i that are misidentified as category j. For the OT dataset, the highest classification rate is 100% for the highway category, and the lowest classification rate is 72% for the open country category. The biggest confusion happens between coast category and open country category. By observing, we find that the misclassified "coast" images show certain similarity to the "open country" images at first glance. While there is no color information to help separate sea water from grassland, the misclassified "coast" images are very easy to be confused with "open country" images. For FP dataset and LS dataset, the biggest confusion happens between the indoor categories (kitchen, living room, and bedroom). By observing the misclassified images, we find that some classification errors are related to the ambiguity of scene images. For example, some "kitchen" images are confused with "living room" images. We find most of them depict the furniture (such as dining table, coffee table, and cabinets) in the central parts of images and the windows in the edge parts of images. They are very easy to be confused. In spite of this, the proposed scheme has achieved good performance. The classification accuracy of three scene datasets is 87.5%, 82.8%, and 78.7%, respectively.

In order to test the influence of different factors (such as kernel functions, scales, and orientations of the Gabor features) on classification performance of the proposed method,



(a) OT dataset

(b) FP dataset

(c) LS dataset

Figure 7: Confusion matrixes of the proposed method.

Table 1: Classification performance of different Gabor features and kernel functions on OT dataset.

Nonlinear kernel	Gabor filter bank		
	4 orientations	8 orientations	12 orientations
RBF			
1 scale	76.82%	77.90%	77.65%
3 scales	78.56%	81.55%	83.06%
5 scales	81.94%	84.22%	82.58%
POLY			
1 scale	73.95%	77.64%	76.78%
3 scales	79.32%	84.36%	82.82%
5 scales	81.16%	83.95%	82.78%
PMK			
1 scale	80.62%	83.46%	83.35%
3 scales	83.98%	84.70%	86.06%
5 scales	85.84%	87.25%	85.42%

Table 2: Classification performance of different Gabor features and kernel functions on FP dataset.

Nonlinear kernel	Gabor filter bank		
	4 orientations	8 orientations	12 orientations
RBF			
1 scale	73.15%	73.56%	73.26%
3 scales	74.14%	78.53%	77.25%
5 scales	77.82%	79.74%	78.16%
POLY			
1 scale	69.42%	73.26%	72.36%
3 scales	74.96%	79.84%	78.44%
5 scales	76.74%	78.55%	78.37%
PMK			
1 scale	76.28%	79.16%	78.86%
3 scales	79.54%	80.48%	81.54%
5 scales	81.42%	82.74%	81.08%

Table 3: Classification performance of different Gabor features and kernel functions on LS dataset.

Nonlinear kernel	Gabor filter bank		
	4 orientations	8 orientations	12 orientations
RBF			
1 scale	67.42%	69.42%	69.28%
3 scales	70.35%	73.68%	74.83%
5 scales	73.56%	75.94%	74.35%
POLY			
1 scale	64.84%	69.35%	68.98%
3 scales	71.15%	76.04%	74.26%
5 scales	72.76%	75.48%	75.32%
PMK			
1 scale	72.38%	74.29%	75.25%
3 scales	75.54%	76.46%	77.65%
5 scales	77.42%	78.85%	77.84%

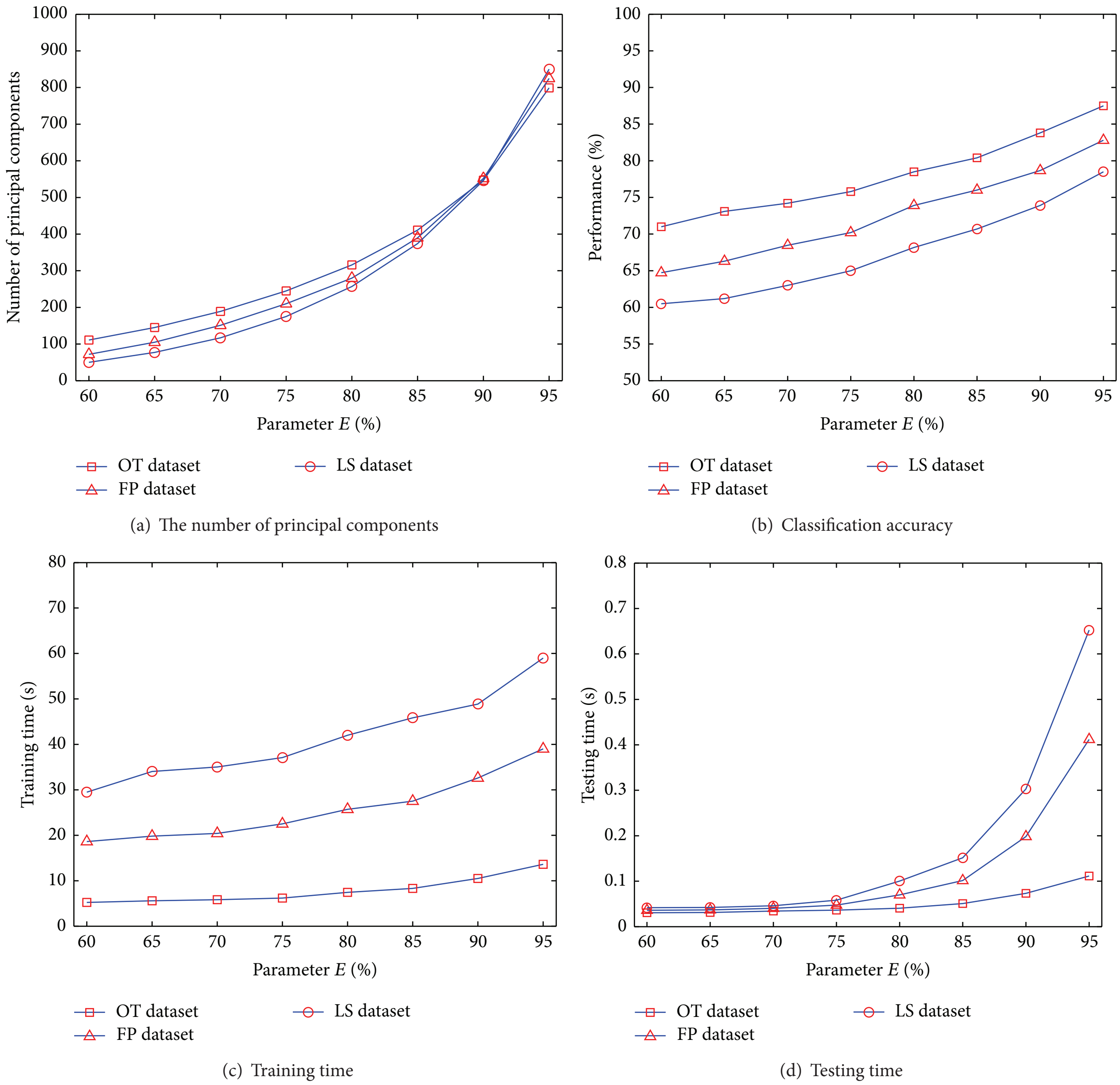
we perform experiments with RBF kernel function, POLY kernel function, and pyramid matching kernel function for three scene datasets, respectively. Tables 1–3 show the performance comparison of these experiments.

In this study, we set the Gaussian width σ of the RBF kernel function to be 1 and set the parameter d of the POLY kernel function $K(x_i, x_j) = [x_i \cdot x_j + 1]^d$ to be 2. As shown in Tables 1–3, the schemes using the RBF kernel function for KPCA obtain better classification performance than the schemes using the POLY kernel function, and the scheme using the PMK for KPCA obtains the highest classification accuracy. The experimental results also show that classification accuracy has an upward trend with the increasing number of directions and scales of extracted Gabor features. But the conclusion cannot be drawn that the more directions and scales of Gabor features are used, the better classification performance is obtained. Owing to the meticulous division, the Gabor features with 12 orientations are not more suitable for scene classification than the Gabor features with 8 orientations. Consequently, the local Gabor features with 5 scales and 8 orientations are the most appropriate for scene classification.

In the proposed method, the nonlinear principal components of feature vectors are extracted by the improved KPCA, and linear "1-a-r" SVMs are used for scene classification. The training time and the testing time decrease relatively owing to dimensionality reduction of feature vectors, and the classification performance changes with the number of retained principal components. Figures 8 and 9 show some experimental curves of our method. The training time and the testing time are the runtime of the linear "1-a-r" SVMs. The experimental environments are given as follows: windows 7, MATLAB7.10, CPU Intel i3-2330M, 2.20 GHz, and 2.00 GB RAM.

Figures 8(a)–8(d) show the experimental curves of the number of principal components, classification accuracy, training time, and testing time when the threshold E changes form 95% to 60% ($T = 95\%$). As shown in Figure 8, the number of principal components declines rapidly when the threshold E decreases. The training time and testing time of "1-a-r" SVMs decrease correspondingly with the reduction of threshold E, and the classification accuracy also decreases correspondingly.

Figures 9(a)–9(d) show the experimental curves of the number of principal components, classification accuracy, training time, and testing time when the threshold T changes form 95% to 60% ($E = 95\%$). Because the principal components with the bigger interclass separability are used for scene classification in our method, good classification performance can be obtained. Figure 9(b) shows the classification accuracy with the various parameter T. Initially, the classification accuracy gradually increases with the decrease of parameter T, because some components with less category information are discarded. After reaching the maximum, the classification accuracy gradually decreases with the decrease of parameter T, because the number of discarded components increases so much that some components with more category information are discarded. The classification accuracy reaches its peak



(a) The number of principal components

(b) Classification accuracy

(c) Training time

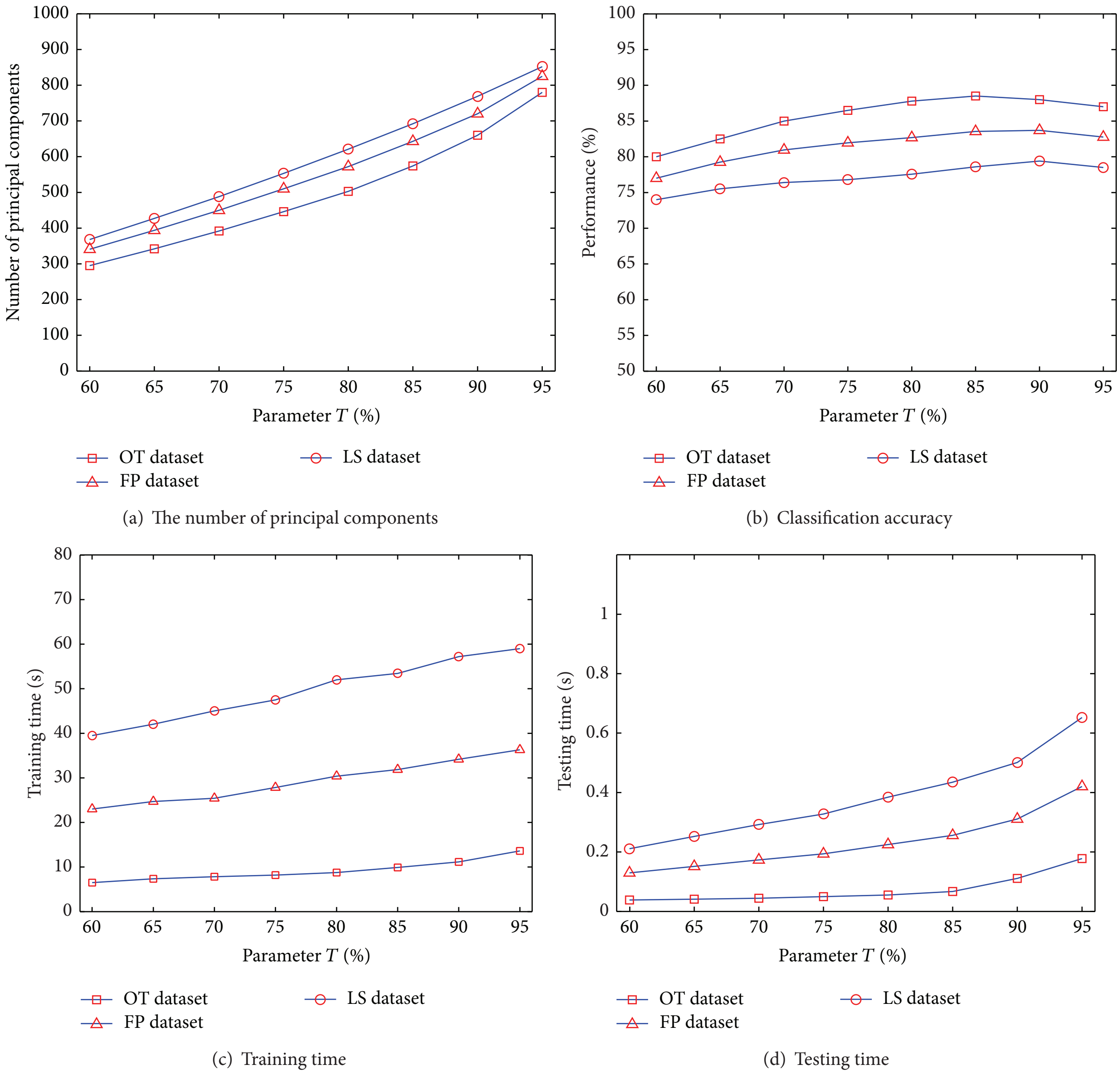
(d) Testing time

Figure 8: Experimental curves ($T = 95\%$).

when T is about 80%–90%. The number of principal components, the training time, and testing time of “1-a-r” SVMs decrease correspondingly with the reduction of threshold T.

In this study, the image size is approximately 300×250. If the images are bigger, the training time and testing time are not affected. The computational cost for local Gabor feature extraction of each scene image is linear with the size of the image. If the images are bigger, the computational cost for feature extraction is higher. However, the training time and the testing time measured in this paper are the runtime of the “1-a-r” SVMs, and therefore the change of the time for feature vector extraction is not included. Moreover, the factors that affect the runtime of SVM classifiers (such as the dimensionality of feature vectors and the number of training images and test images) are not related to the size of the image. Even if the images are bigger, the runtime of “1-a-r” SVMs is unchanged.

The proposed method is also compared with several well-known algorithms, such as the dense SIFT method [11], the BOVW method [4], and the “Gist” method [14]. We randomly select 200 scene images of each category from three different datasets as experiment images. Half of them are used as training samples and the others are used for testing. The penalty factor C of “1-a-r” SVMs is set to be 10. In the dense SIFT method, the sampling interval of the dense regular grid is 8 pixels. SIFT descriptors are computed from $16 * 16$ image patches. The vocabulary size is 300, and the number of levels of the spatial pyramid is three. The other



(a) The number of principal components

(b) Classification accuracy

(c) Training time

(d) Testing time

Figure 9: Experimental curves ($E = 95\%$).

parameter settings are the same as the settings in [11]. “1-a-r” SVMs with the spatial pyramid matching kernel are used for scene classification. In the BOVW method, Difference of Gaussian (DoG) detectors are used to automatically detect key points. SIFT descriptors are adopted for representing local features of scene images, and “1-a-r” RBF-SVMs are used for scene classification. We set the Gaussian width σ of the RBF kernel function to be 1. The other parameter settings are the same as the settings in [4]. In the “Gist” method, “Gist” feature is extracted from a $4 * 4$ grid of the filtering output of a scene image convolved with 40 Gabor filters (5 scales and 8 orientations), which have been described in Section 2. “1-a-r” SVMs with the RBF kernel function are used for scene classification. The Gaussian width σ of the RBF kernel function is set to be 1.

Figure 10 shows the classification accuracy of different methods. For three different scene datasets, the proposed method is slightly better than the dense SIFT method and much better than the BOVW method and the “Gist” method. In the proposed method, local Gabor features extracted by imitating the “Gist” model, which conforms to the mechanism of human vision, have good discrimination abilities for sampling patches of scene images. So the accuracy of visual words which are obtained by the K-means clustering algorithm can be guaranteed. Meanwhile the improved KPCA is used for extracting nonlinear principal components. The principal components containing more category information, which are suitable for scene classification, are retained. The proposed method achieves considerably higher accuracy.

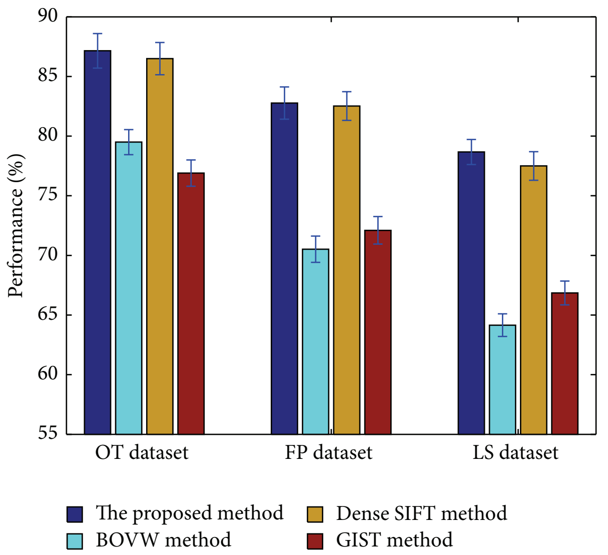


FIGURE 10: Performance comparison of different classification methods.

4. Conclusions

A new scene classification method has been proposed based on local Gabor features. The local Gabor feature descriptors which are extracted according to the "Gist" theory have sufficient discrimination abilities for sampling patches of scene images. By quantizing local Gabor features into discrete codewords and employing a spatial pyramid matching model, pyramid histograms of visual words which contain spatial position information of images are obtained for representing scene images. In addition, the principal components of PHOW containing more category information are extracted by an improved KPCA method. These principal components are suitable for scene classification, and they can improve both classification accuracy and computational cost. Numerical experiments are conducted on three scene datasets. The experimental results demonstrate the effectiveness of the method. The proposed method can also be extended to different applications such as the classification of commodity images and the classification of event images.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported by the Science Research Foundation of the Education Department of Liaoning Province of China (Grant no. L2014174).

References

[1] X. Zhou, X. D. Zhuang, H. Tang, M. Hasegawa-Johnson, and T. S. Huang, "Novel Gaussianized vector representation for improved natural scene categorization," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 702–708, 2010.

[2] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image classification for content-based indexing," *IEEE Transactions on Image Processing*, vol. 10, no. 1, pp. 117–130, 2001.

[3] N. Serrano, A. E. Savakis, and J. Luo, "Improved scene classification using efficient low-level features and semantic cues," *Pattern Recognition*, vol. 37, no. 9, pp. 1773–1784, 2004.

[4] F.-F. Li and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 524–531, June 2005.

[5] P. Quelhas, F. Monay, J.-M. Odobez, D. Gatica-Perez, and T. Tuytelaars, "A thousand words in a scene," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1575–1589, 2007.

[6] L. Nanni, A. Lumini, and S. Brahnam, "Ensemble of different local descriptors, codebook generation methods and subwindow configurations for building a reliable computer vision system," *Journal of King Saud University—Science*, vol. 26, no. 2, pp. 89–100, 2014.

[7] Z. Li and K.-H. Yap, "An efficient approach for scene categorization based on discriminative codebook learning in bag-of-words framework," *Image and Vision Computing*, vol. 31, no. 10, pp. 748–755, 2013.

[8] J. Qin and N. H. C. Yung, "Scene categorization via contextual visual words," *Pattern Recognition*, vol. 43, no. 5, pp. 1874–1888, 2010.

[9] N. M. Elfiky, J. Gonzàlez, and F. X. Roca, "Compact and adaptive spatial pyramids for scene recognition," *Image and Vision Computing*, vol. 30, no. 8, pp. 492–500, 2012.

[10] L. Zhou, Z. T. Zhou, and D. W. Hu, "Scene classification using a multi-resolution bag-of-features model," *Pattern Recognition*, vol. 46, no. 1, pp. 424–433, 2013.

[11] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2169–2178, June 2006.

[12] A. Oliva and A. Torralba, "Chapter 2 Building the gist of a scene: the role of global image features in recognition," *Progress in Brain Research*, vol. 155, pp. 23–36, 2006.

[13] F. F. Li, R. VanRullen, C. Koch, and P. Perona, "Rapid natural scene categorization in the near absence of attention," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 14, pp. 9596–9601, 2002.

[14] A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[15] K. Hotta, "Local co-occurrence features in subspace obtained by KPCA of local blob visual words for scene classification," *Pattern Recognition*, vol. 45, no. 10, pp. 3687–3694, 2012.

[16] K. Hotta, "Local autocorrelation of similarities with subspaces for shift invariant scene classification," *Pattern Recognition*, vol. 44, no. 4, pp. 794–799, 2011.

[17] A. Bolovinou, I. Pratikakis, and S. Perantonis, "Bag of spatio-visual words for context inference in scene classification," *Pattern Recognition*, vol. 46, no. 3, pp. 1039–1053, 2013.

[18] L. Nanni and A. Lumini, "Heterogeneous bag-of-features for object/scene recognition," *Applied Soft Computing Journal*, vol. 13, no. 4, pp. 2171–2178, 2013.

[19] X. L. Meng, Z. Z. Wang, and L. Z. Wu, “Building global image features for scene recognition,” *Pattern Recognition*, vol. 45, no. 1, pp. 373–380, 2012.

[20] J. Li, X. Li, and D. Tao, “KPCA for semantic object extraction in images,” *Pattern Recognition*, vol. 41, no. 10, pp. 3244–3250, 2008.

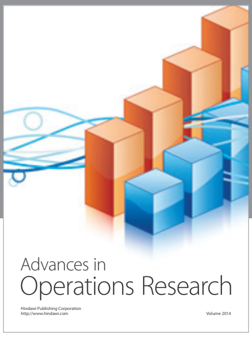
[21] P. F. Jia, F. C. Tian, Q. H. He, S. Fan, J. L. Liu, and S. X. Yang, “Feature extraction of wound infection data for electronic nose based on a novel weighted KPCA,” *Sensors and Actuators B: Chemical*, vol. 201, pp. 555–566, 2014.

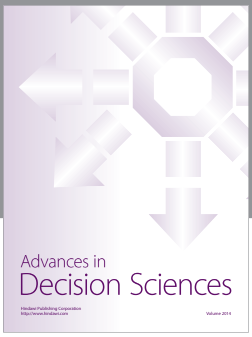
[22] Y. W. Zhang, “Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM,” *Chemical Engineering Science*, vol. 64, no. 5, pp. 801–811, 2009.

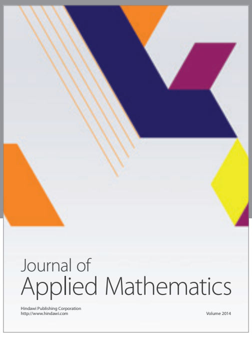
[23] M. X. Jia, H. Y. Xu, X. F. Liu, and N. Wang, “The optimization of the kind and parameters of kernel function in KPCA for process monitoring,” *Computers and Chemical Engineering*, vol. 46, pp. 94–104, 2012.

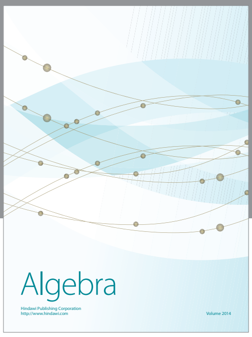
[24] Y. Xu, D. Zhang, F. Song, J.-Y. Yang, Z. Jing, and M. Li, “A method for speeding up feature extraction based on KPCA,” *Neurocomputing*, vol. 70, no. 4–6, pp. 1056–1061, 2007.

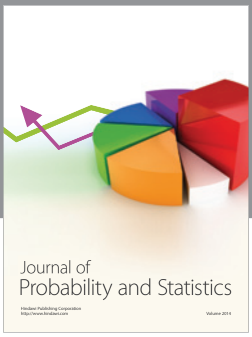
[25] C.-W. Hsu and C.-J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.











Hindawi

Submit your manuscripts at
http://www.hindawi.com