



Hierarchical feature concatenation-based kernel sparse representations for image categorization

Bo Wang¹ · Jichang Guo¹ · Yan Zhang¹ · Chongyi Li¹

© Springer-Verlag Berlin Heidelberg 2016

Abstract In order to obtain improved performance in complicated visual categorization tasks, considerable research has adopted multiple kernel learning based on dozens of different features. However, it is a complex process that needs to extract a multitude of features and seeks the optimal combination of multiple kernels. Inspired by the key idea of hierarchical learning, in this paper, we propose to find sparse representation based on feature concatenation using hierarchical kernel orthogonal matching pursuit (HKOMP). In addition to commonly used spatial pyramid feature for kernel representation, our method only employs one type of generic image feature, i.e., p.d.f gradient-based orientation histogram for concatenation of sparse codes. Next, the resulting concatenated features kernelized with widely used Gaussian radial basis kernel function form compact sparse representations in the second layer for linear support vector machine. HKOMP algorithm combines the advantages of building image representations layer-by-layer and kernel learning. Several publicly available image datasets are used to evaluate the presented approach and empirical results for various datasets show that the proposed scheme outperforms many kernel learning based and other competitive image categorization algorithms.

Keywords Image categorization · Kernel sparse representation · Feature concatenation · Hierarchical learning · Linear support vector machine

1 Introduction

Image categorization has attracted close attentions in the research area of computer vision over the past decade [1–3], because categorization task can serve as a building block for other tasks comprising image search, object retrieval, location recognition, autonomous robotics, etc. [4–9]. Since sparse representation has proven to be a powerful tool for signal processing, it has been widely applied to a great variety of problems in computer vision, including image denoising [10], image annotation [11], image and video categorization [12], visual tracking [13], and so on. The success of sparse representation is mainly due to three factors. Firstly, the characteristics of sparseness are ubiquitous in natural image patches. Secondly, sparse representations are more likely to be separable in high-dimensional spaces. Last but not least, sparse coding can achieve a much lower reconstruction error which is useful to deal with the noise existing in images.

However, most existing techniques perform sparse coding step directly in image feature space. Since the data in the same direction would be entangled after the normalization process, the data in the same direction from different classes are not linearly separable in this space [14]. Moreover, the images are nonlinearly sparse and various descriptors in computer vision have nonlinear similarity measurement. Therefore, kernel trick based methods mapping the nonlinear features into a high-dimensional feature space have been presented in recent years [15–19]. The possibly infinite-dimensional feature space which is a reproducing kernel Hilbert space (RKHS) can be defined by Mercer kernel satisfying Mercer's condition [20]. Essentially, the Mercer kernel is a valid function as long as it is symmetric, continuous and positive semi-definite. As a result, building linear models in its feature space can provide the power of corresponding nonlinear models in the input space.

✉ Bo Wang
neuwb@tju.edu.cn

¹ School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China

On the other hand, while many efforts have been devoted to design discriminative and robust features for image categorization over the past few years [21–24], in practice it has been well acknowledged that single type of derived feature could not provide high discrimination for all classes. Hence, a number of algorithms combine a set of diverse and complementary features comprising appearance, shape, color, texture, semantics information and the advantage of combination has been demonstrated in recent literature [25–31]. In order to overcome the problems of intra-class variability and inter-class correlation, multiple kernel learning (MKL) [32] which fuses multiple image features in a unified kernel space instead of single kernel has become popular for intricate visual tasks. In other words, the purpose of MKL is to provide an effective scheme of fusing features and the associated kernels and thereby determine the mixing weights of different features from training data.

Intuitively, scores of complementary features are more likely to represent different attributes of a great variety of image classes and thus result in a more powerful classifier than any individual feature; whereas, extraction of an army of diverse features and selection of corresponding distance measurement are complicated and time consuming, even require adequate experience of choice. For instance, local binary patterns (LBP) [33] and its variant called average local binary pattern (ALBP) [34] can be regarded as useful texture features based on occurrence histogram of local binary patterns. Likewise, image regions with similar local structures and texture can be described by similar covariance matrices, i.e., region covariance descriptors [35]. While scale-invariant feature transform (SIFT) [36] descriptors adept in finding repeated image content are exploited to imply appearance information, typically, pyramid of histograms of orientation gradients (PHOG) [37] and self-similarity descriptors (SSIM) [22] are employed as shape features. Furthermore, GIST features are presented specifically for scene recognition [38]. As for color features, color histograms built in different color spaces including RGB, HSV and LAB are often adopted for object recognition. According to different characteristics of the datasets, some algorithms can also

choose other types of features to improve performance, such as V1S+ [39] and geometric blur [40].

In addition to a wide range of features themselves, appropriate distance functions used to construct the kernel matrices also play a vital role in kernel learning and should be further studied in the future. For example, χ^2 distance is always employed for LBP, PHOG, SSIM and color histograms, while geodesic distance and RBF kernel with the average distance are generally used for the region covariance descriptors and geometric blur features, respectively [41]. Unfortunately, significant research efforts in improving MKL often provide conflicting results in terms of efficiency and effectiveness [42], and it is observed that even sophisticated algorithms cannot clearly outperform the simple average of kernels [43].

To overcome the above-mentioned issues, this paper proposes a hierarchical feature concatenation method using efficient kernel KSVD (KKSVD) for dictionary learning and KOMP for sparse coding. More specifically, the creating kernel sparse representations (KSR) utilizing p.d.f gradient-based orientation histogram [44] and dense SIFT in spatial pyramid [45], respectively, are concatenated as the input features for the second layer. Subsequently, the concatenated features form resulting compact KSR based on KKSVD and KOMP again in the second layer. Accordingly, high complexity and even ambiguity involved in MKL can be avoided. Besides, the final image representations have relatively low-dimensional structure which only depends on the number of training examples and total categories in datasets. Experimental results demonstrate that the proposed algorithm significantly outperforms some kernel learning based approaches and other state-of-the-art methods on several publicly available datasets, i.e., UIUC-Sports, Scene15, Oxford Flowers and Caltech101. Figure 1 presents an overview of our scheme.

The major contributions of this paper are as follows:

1. We study the performance of the p.d.f gradient-based orientation histogram feature in HKOMP model for image categorization. Despite the original work makes the dimensionality of feature very high, it is effectively

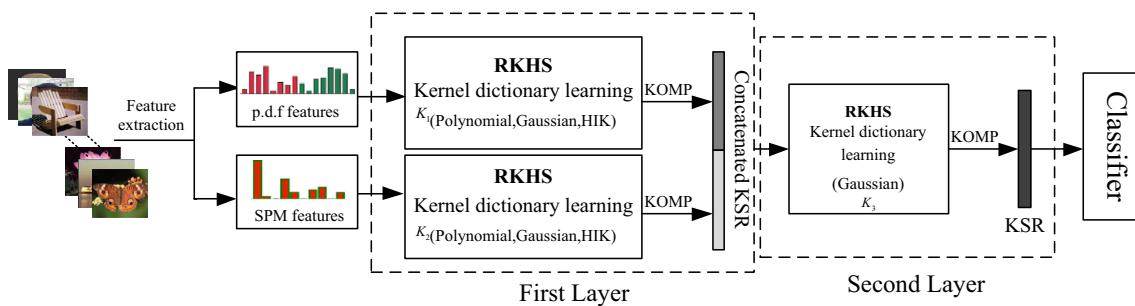


Fig. 1 Overview of the proposed method

- worth applying for kernel learning tasks due to its discriminative power for any kinds of tasks.
2. The proposed algorithm only exploits two types of features in HKOMP and the dense SIFT in spatial pyramid as complementary one shows surprising results after feature concatenation.
 3. The resulting KSR have much lower dimensionality than that of other patterns, which is beneficial for feature concatenation in terms of efficiency.

The remaining part of this paper is organized as follows: in Sect. 2, we briefly review some related works. We provide necessary preliminaries for our proposed method in Sect. 3. In Sect. 4, a novel batch kernel orthogonal matching pursuit (BKOMP) is developed for KSRC. Some experimental results and analysis are reported in Sect. 5. Section 6 concludes this paper.

2 Related works

In the past decade, kernel-based learning methods have shown a great potential in many real world applications. Kernel sparse representation-based classification (KSRC) [14, 18, 19, 46], as one of the most important tasks, is introduced to overcome the inherent deficiency of sparse representation-based classification (SRC) and obtains improved performance. In essence, these approaches map the nonlinear data into high-dimensional feature space employing the kernel trick with the purpose that the same distribution are readily grouped together and are linearly separable in this space. It should be noted that KSRC suggested in [14] is different from the other three methods since it incorporates the dimensionality reduction in the kernel feature space with KSRC.

However, the methods mentioned above achieve KSR via convex relaxations using l_1 -norm which can be replaced by a greedy fashion, i.e., orthogonal matching pursuit (OMP). Since OMP performs extremely fast in real-time applications, [47] investigates the KOMP and modifies it into single-step KOMP (S-KOMP) for low-dimensional face recognition, while [48] proposes a nonlinear kernel dictionary learning method and once sparse codes are found employing KOMP in the feature space, the dictionary atoms will be updated in an efficient way. As an extension of [17], [48] presents an effective system that combines multiple classifiers based on the kernel sparse codes for object recognition.

On the other hand, the idea of hierarchical learning has been successfully applied into many models for image classification. Typically, [49–52] show that good image representations which guarantee promising performance can be derived from pixel level via hierarchical structure, whereas these studies consider performing sparse coding directly in the image feature space rather than RKHS.

As to image feature extraction, previous research has made considerable contributions to diverse discriminative features over the past few years. A seminal approach to extract effective features for image classification is proposed in [44]. Stemming from bag-of-features model that needs to extract a plenty of local descriptors from an image, the proposed scheme is built upon the probability density function (p.d.f) obtained by applying kernel density estimator to those local descriptors. The orientations of gradients on p.d.f which are subsequently coded and aggregated into the histograms effectively characterize the shape of the p.d.f from the geometrical perspective. Most importantly, the features are so general as to be applicable to various datasets for classification tasks. Besides, the traditional spatial pyramid matching (SPM) [45] can also provide generic image features which are critical for success.

It should be mentioned that our algorithm is different from [44] which employs sequential minimal optimization (SMO) algorithm for SVM. While [17] utilizes kernel dictionary learning as well, it combines 39 different features using boosting method which has to seek correct weighting of multiple features. To the best of our knowledge, our algorithm is the first one which incorporates the hierarchical KSR using feature concatenation.

3 Preliminaries

The goal of sparse coding is to represent a signal by the linear combination of a small number of atoms chosen out of a dictionary. In particular, the dictionary learning from the training data is more effective to model the complex local structures of natural images. Given a set of N_c images in the c th class, the dictionary D and sparse codes X can be found via the following optimization problem,

$$\arg \min_{D, X} \|Y - DX\|_F^2 \quad \text{s.t.} \quad \|x_i\|_0 \leq T, \quad \forall i = 1 \dots N_c, \quad (1)$$

where $Y = [y_1, \dots, y_{N_c}]$, $Y \in R^{n \times N_c}$, $D \in R^{n \times K}$, $X = [x_1, \dots, x_{N_c}]$, $X \in R^{K \times N_c}$, the notation $\|\cdot\|_F$ denotes the Frobenius norm, x_i represents the i th column of X , the zero-norm $\|\cdot\|_0$ counts the nonzero entries in the sparse code x_i , T is the sparsity level which bounds the number of nonzero elements.

Although some techniques are able to handle this problem efficiently, e.g., KSVD, they fail to capture nonlinear structures of the data. Therefore, kernel methods mapping the data into a higher dimensional feature space have been introduced into KSVD algorithm to overcome the limitation. Typically, as stated in [48], it employs the multiplication of the nonlinear mapping of input data and a transformation matrix to take the place of the original dictionary D .

Let $\phi : \mathbb{R}^n \rightarrow \mathbf{H}$ be a nonlinear mapping that transforms the features to a unique associated RKHS. In this feature space, the signals $Y = [y_1, \dots, y_{N_c}]$ can now be represented as $\phi(Y) = [\phi(y_1), \dots, \phi(y_{N_c})]$. Since the dimension of the feature space can be infinitely large, the optimization problem can be reformulated as follows:

$$\arg \min_{Z, X} \|\phi(Y) - \phi(Y)ZX\|_F^2 \quad \text{s.t.} \quad \|x_i\|_0 \leq T, \\ \forall i = 1 \dots N_c \quad (2)$$

Then, denote the objective function as follows:

$$\begin{aligned} \|\phi(Y) - \phi(Y)ZX\|_F^2 &= \|\phi(Y)(I - ZX)\|_F^2 \\ &= \text{trace}((I - ZX)^T \phi(Y)^T \phi(Y)(I - ZX)) \\ &= \text{trace}((I - ZX)^T K(Y, Y)(I - ZX)), \end{aligned} \quad (3)$$

where I and Z are the identity and transformation matrix, respectively. $K(Y, Y)$ is referred to as kernel matrix whose elements are computed from $\kappa(y_i, y_j) = \langle \phi(y_i), \phi(y_j) \rangle$. The replacement of inner product by the kernel function $\kappa(y_i, y_j)$ amounts to a conversion mapping the nonlinear problem in the input space into a linear one in the feature space. In other words, kernel functions can implicitly specify the mapping ϕ without having to know the nonlinear feature mapping function explicitly. Several widely used kernel functions in practice are the linear, polynomial and Gaussian radial basis function (GRBF) kernels, respectively, which can be listed as follows:

(i) Linear:

$$\kappa(x, y) = \langle x, y \rangle \quad (4)$$

(ii) Polynomial:

$$\kappa(x, y) = (\langle x, y \rangle + a)^b, \quad b > 0, \quad a \in R \quad (5)$$

(iii) Gaussian radial basis function (GRBF):

$$\kappa(x, y) = \exp\left(-\frac{\|x - y\|^2}{d}\right), \quad d > 0 \quad (6)$$

Furthermore, the histogram intersection kernel (HIK) has been demonstrated to provide improved performance for a

variety of tasks when histogram features are exploited [53]. Since HIK can generate a positive definite kernel which facilitates its use in classifiers, it is worth introducing HIK into our experiments. Let $\mathbf{h} = (h_1, \dots, h_t) \in R_+^t$ be a histogram, HIK can be defined as follows:

$$\kappa_{\text{HI}}(\mathbf{h}_1, \mathbf{h}_2) = \sum_{q=1}^t \min(h_{1q}, h_{2q}), \quad (7)$$

where \mathbf{h} could signify SIFT descriptors or histogram of code words.

3.1 Orthogonal matching pursuit (OMP) and kernel OMP (KOMP)

It is well known that OMP is one of the greedy algorithms for sparse approximation due to its simplicity and efficiency. Since the optimization problem (1) can be solved in an alternating fashion, OMP is capable of computing sparse codes when this problem is decoupled to N simpler sub-problems of the following form in the first stage:

$$\arg \min_{x_i} \|y_i - D x_i\|_2^2 \quad \text{s.t.} \quad \|x_i\|_0 \leq T \quad (8)$$

The pipeline of OMP is shown in Algorithm 1. It is easy to find that the residual can be obtained as the following form.

$$\begin{aligned} \|r_s\|_2^2 &= (y - A_s x_s)^T (y - A_s x_s) \\ &= y^T y - 2y^T A_s x_s + x_s^T A_s^T A_s x_s \end{aligned} \quad (9)$$

Secondly, the correlations between r_{s-1} and d_j are computed as the similar form,

$$\begin{aligned} \langle r_{s-1}, d_j \rangle &= \langle y - A_{s-1} x_{s-1}, d_j \rangle \\ &= y^T d_j - A_{s-1}^T x_{s-1}^T d_j \end{aligned} \quad (10)$$

Most importantly, it is observed that the solution of the least square problem can also be expressed in the form of inner products as follows,

$$x_s = (A_s^T A_s)^{-1} A_s^T y \quad (11)$$

Consequently, OMP is naturally connected with kernel trick according to all the steps in Algorithm 1.

 Algorithm 1: Orthogonal Matching Pursuit (OMP)

1. Input: Signal $y \in R^n$, Dictionary $D = [d_1, \dots, d_K] \in R^{n \times K}$, Sparsity level T .
 2. Output: Sparse code $x \in R^K$ such that $y \approx Dx$.
 3. Initialization: Residual $r_0 = y$, Index set $\beta_0 = \emptyset$, Selected atom set $A_0 = \emptyset$.
 4. For $s = 1 : T$
 5. $\alpha_s = \arg \max_{j=1, \dots, K} \langle r_{s-1}, d_j \rangle$
 6. $\beta_s = [\beta_{s-1} \quad \alpha_s]$
 7. $A_s = [A_{s-1} \quad d_{\alpha_s}]$
 8. Approximate the coefficients of the selected atoms by least squares:

$$x_s = \arg \min_x \|y - A_s x\|_2$$
 9. Compute the new residual:

$$r_s = y - A_s x_s$$
 10. End
-

As a result, the new formulation of the sub-problem based on (2) can be represented by:

$$\arg \min_{x_i} \|\phi(y_i) - \phi(Y)Zx_i\|_2^2 \quad \text{s.t.} \quad \|x_i\| \leq T \quad (12)$$

According to the analysis above, the pipeline of OMP in high-dimensional space is shown in Algorithm 2. Note that the solution of the least square problem here can be rewritten as:

$$\begin{aligned} x_s &= ((\phi(Y)Z\beta)^T \phi(Y)Z\beta)^{-1} (\phi(Y)Z\beta)^T \phi(y) \\ &= \left(Z_\beta^T \phi(Y)^T \phi(Y) Z_\beta \right)^{-1} Z_\beta^T \phi(Y)^T \phi(y) \\ &= \left(Z_\beta^T K(Y, Y) Z_\beta \right)^{-1} (K(y, Y) Z_\beta)^T, \end{aligned} \quad (13)$$

where Z_β indicates the selected atom set whose indices are from the set β_s . Hence, KOMP achieves sparse representation in any high-dimensional RKHS which is implicitly determined by kernel function as long as it satisfies the Mercer's condition.

3.2 Kernel KSVD (KKSVD)

Over the past decade, KSVD as an efficient iterative method has been widely used for dictionary learning and has shown good performance in many tasks, whereas KSVD cannot be applicable to nonlinear models which arise in practical

applications. Therefore, kernelized KSVD proposed in [17] addresses this problem in kernel space. Let the objective function in (2) be rewritten as follows:

$$\begin{aligned} \|\phi(Y) - \phi(Y)ZX\|_F^2 &= \left\| \phi(Y) - \phi(Y) \sum_{g=1}^K z_g x_T^g \right\|_F^2 \\ &= \left\| \phi(Y) \left(I - \sum_{g \neq k} z_g x_T^g \right) - \phi(Y) z_k x_T^k \right\|_F^2, \end{aligned} \quad (14)$$

where z_k and x_T^g denote the k th column of Z and the g th row of X , respectively. It should be noted that the first term of (14) represents the error between the true and approximated signals when removing the k th atoms. The second term of (14) represents the contribution of k th atom to approximated signals. The pipeline of KKSVD is briefly shown in Algorithm 3.

4 Batch kernel orthogonal matching pursuit (BKOMP) and feature concatenation for KSRC

The total cost of KOMP can be reduced by batch kernel orthogonal matching pursuit (BKOMP) in practice. As can be seen in Algorithm 4, the key step is atom selection which is most time-consuming.

 Algorithm 2: Kernel Orthogonal Matching Pursuit (KOMP)

1. Input: Signal $y \in R^n$, $Y \in R^{n \times N_c}$, Transformation matrix $Z \in R^{N_c \times K}$
Sparsity level T , Kernel function κ , Kernel matrix K .
 2. Output: Sparse code $x \in R^K$ such that $\phi(y) \approx \phi(Y)Zx$.
 3. Initialization: Index set $\beta_0 = \emptyset$, Current loading of the signal $\phi(y)$ over the base $\phi(Y)$,
 $P_0 = Zx_0 = 0$
 4. For $s = 1 : T$
 5. $\alpha_s = \arg \max_{j=1,\dots,K} (K(y, Y) - P_s^T K(Y, Y))z_j \quad \forall j \notin \beta_{s-1}$
 6. $\beta_s = [\beta_{s-1} \quad \alpha_s]$
 7. Approximate the coefficients of the selected atoms by least squares:
 $x_s = (Z_\beta^T K(Y, Y)Z_\beta)^{-1}(K(y, Y)Z_\beta)^T$
 8. Compute the new loading of the signal $\phi(y)$ over the base $\phi(Y)$:
 $P_s = Z_\beta x_s$
 9. End
-

 Algorithm 3: Kernel KSVD (KKSVD)

1. Input: Signal $Y \in R^{n \times N_c}$, Transformation matrix $Z \in R^{N_c \times K}$.
Kernel function κ , Kernel matrix K .
 2. Output: Transformation and sparse codes matrix Z and X .
 3. Initialization: Set T random elements of each column in X to be 1. Set iteration $it = 1$.
 4. 1st stage (sparse coding): Given a fixed transformation matrix Z , utilize the KOMP algorithm described in Algorithm 2 to obtain sparse codes matrix X .
 5. 2nd stage (dictionary update): Update each column in $Z^{(it)}$ by applying SVD decomposition.
 6. Set $it = it + 1$
-

Where D^r are the rows of D , G_Ω and $G_{\Omega\Omega}$ denote the sub-matrices of G containing the columns and both rows and the columns indexed by Ω . When collective approach which has been proven more useful than distributive approach is adopted in our scheme, the final dictionary from all classes is computed by:

$$D = [D_1 \dots D_{N_c}] = [\phi(Y_1)Z_1 \dots \phi(Y_{N_c})Z_{N_c}] \quad (15)$$

According to the fastest implementation of BOMP [50], the projections $D^T \phi(Y_o)$ can be rewritten as follows:

$$\begin{aligned} D^T \phi(Y_o) &= [\phi(Y_1)Z_1 \dots \phi(Y_{N_c})Z_{N_c}]^T \phi(Y_o) \\ &= \begin{bmatrix} Z_1^T K(Y_1, Y_o) \\ Z_2^T K(Y_2, Y_o) \\ \vdots \\ Z_{N_c}^T K(Y_{N_c}, Y_o) \end{bmatrix} \end{aligned} \quad (16)$$

Algorithm 4: batch kernel orthogonal matching pursuit (BKOMP)

1. Input: Dictionary D , observation $\phi(Y_o)$, Sparsity level T .
2. Output: Sparse coefficients matrix X .
3. Initialization: $\Omega = \emptyset$, $\eta^0 = D^T \phi(Y_o)$, $G = D^T D$.
4. For $w = 1:T$
5. Selecting the new atom: $\bar{w} = \arg \max_w |\eta_w|$
6. $\Omega = \Omega \cup \bar{w}$
7. Updating the sparse coefficients: $x_\Omega = G_{\Omega\Omega}^{-1} \eta_\Omega^0$
8. Updating η : $\eta = \eta^0 - G_{\Omega\Omega} x_\Omega$
9. End

The corresponding Gram matrix $D^T D$ using final dictionary with normalized columns can be represented as:

$$D^T D = [\phi(Y_1)Z_1 \dots \phi(Y_{N_c})Z_{N_c}]^T [\phi(Y_1)Z_1 \dots \phi(Y_{N_c})Z_{N_c}] \\ = \begin{bmatrix} Z_1^T K(Y_1, Y_1)Z_1 & Z_1^T K(Y_1, Y_2)Z_2 & \dots & Z_1^T K(Y_1, Y_{N_c})Z_{N_c} \\ Z_2^T K(Y_2, Y_1)Z_1 & Z_2^T K(Y_2, Y_2)Z_2 & \dots & Z_2^T K(Y_2, Y_{N_c})Z_{N_c} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{N_c}^T K(Y_{N_c}, Y_1)Z_1 & Z_{N_c}^T K(Y_{N_c}, Y_2)Z_2 & \dots & Z_{N_c}^T K(Y_{N_c}, Y_{N_c})Z_{N_c} \end{bmatrix} \quad (17)$$

Inspired by kernelizing KSVD, the p.d.f gradient-based orientation histograms which have been demonstrated as powerful tools for image categorization can be used to extract more generic features in kernel methods. The resulting KSR using KKSVD and KOMP show compact low-dimensional characteristics, which makes it possible to employ feature concatenation for hierarchical learning.

In order to obtain more discriminative information, SPM features which are commonly used in kernel methods can be combined with kernelized p.d.f features in the first layer. Therefore, only two types of kernelized features are concatenated as input for the second layer. The concatenated KSR are not only very compact, but also can be captured the nonlinear structures with kernelized KSVD.

For instance, the number of training examples in Caltech101 is typically set to be 30 in comparison with other experiments. Besides, this dataset comprised 101 classes and an additional background class. As a result, the dimension of concatenated features which depends on the product of the number of classes and training examples is no more than

Table 1 Dimensions of concatenated features in different datasets

| Datasets | Maximum number of training examples | Number of classes | Maximum dimension of KSR |
|----------------|-------------------------------------|-------------------|--------------------------|
| UIUC-Sports | 70 | 8 | 1120 |
| Scene15 | 100 | 15 | 3000 |
| Oxford Flowers | 40 | 17 | 1360 |
| Caltech101 | 30 | 102 | 6120 |

Table 2 Maximum dimensions and memory usage of the matrices of training data in different datasets (second layer)

| Datasets | Maximum dimensions of the matrix of training data | Memory usage (double) |
|----------------|---|-----------------------|
| UIUC-Sports | 1120×560 | 0.005 GB |
| Scene15 | 3000×1500 | 0.034 GB |
| Oxford Flowers | 1360×680 | 0.007 GB |
| Caltech101 | 6120×3060 | 0.14 GB |

6120. For convenience, Table 1 summarizes the dimension of concatenated features in different datasets exploited in our experiments.

Hence, the matrix of training and testing data employed for the second layer has much smaller size which is beneficial for any types of matrix operations. The advantage of our proposed scheme is fit for hierarchical learning in terms of efficiency. Table 2 provides the maximum dimensions and memory usage of the matrix of training data in the second layer on different datasets.

5 Experimental results

In this section, we apply hierarchical feature concatenation to KSRC on four widely used datasets, i.e., UIUC-Sports, Scene15, Oxford Flowers and Caltech101, which are evaluated for event, scene and object classification tasks, respectively.

The UIUC-sports dataset contains 8 sport categories, e.g., polo, bocce, snowboarding, rock climbing. The total number of images is 1579 and 137–250 in each class. We follow the standard experimental setup that 70 images for training and 60 for testing at random per category. This database is arguably quite challenging due to variations of sizes and poses across over each category with cluttered background. The sample images are visualized in Fig. 2.

The Scene15 can be regarded as an statistic scene category dataset which consists of 15 categories of 4485 indoor and outdoor images. Each category contains 200–400 images, such as kitchen and bedroom for indoor, forest and coast for outdoor. As for this dataset, we randomly choose 100 images from each class as training examples and use the left for testing. Figure 3 shows some sample images of Scene15 dataset.

Fig. 2 Sample images for the UIUC-sports dataset

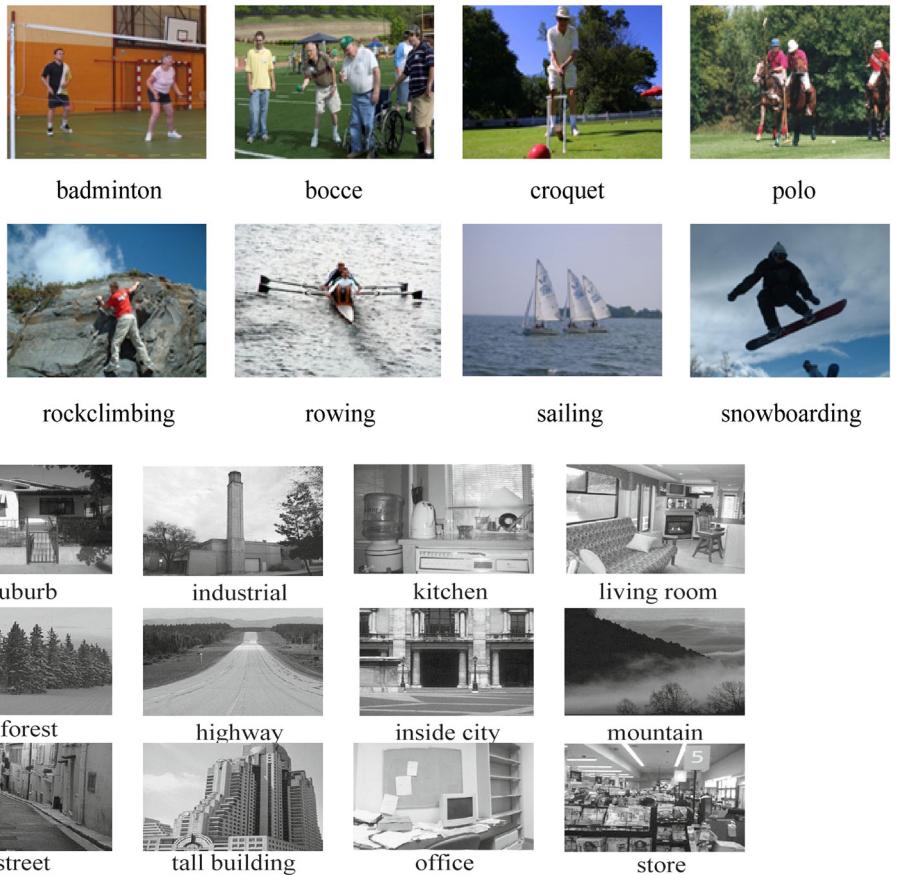
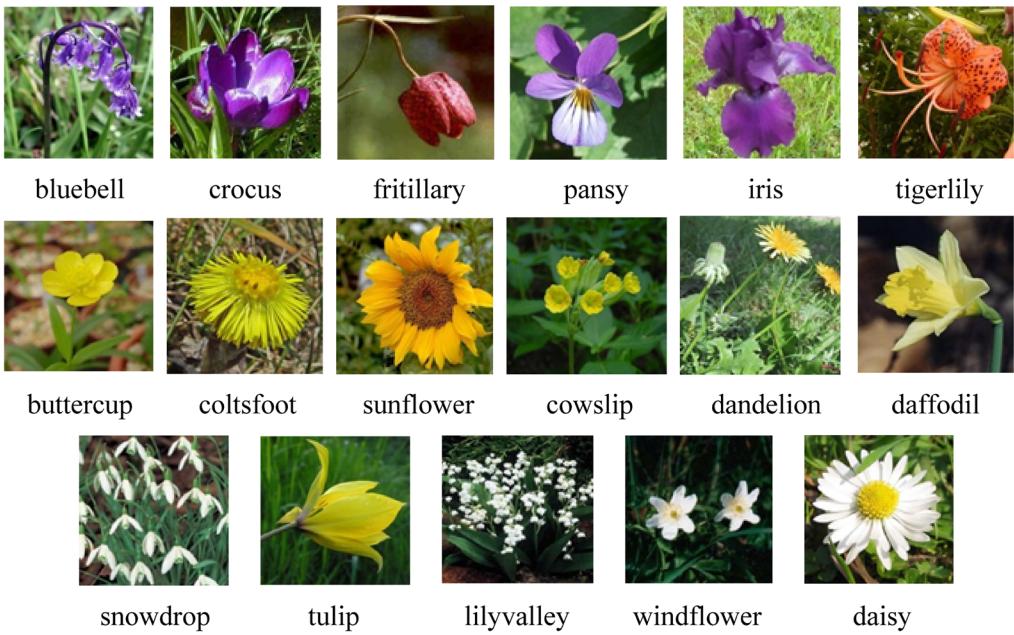
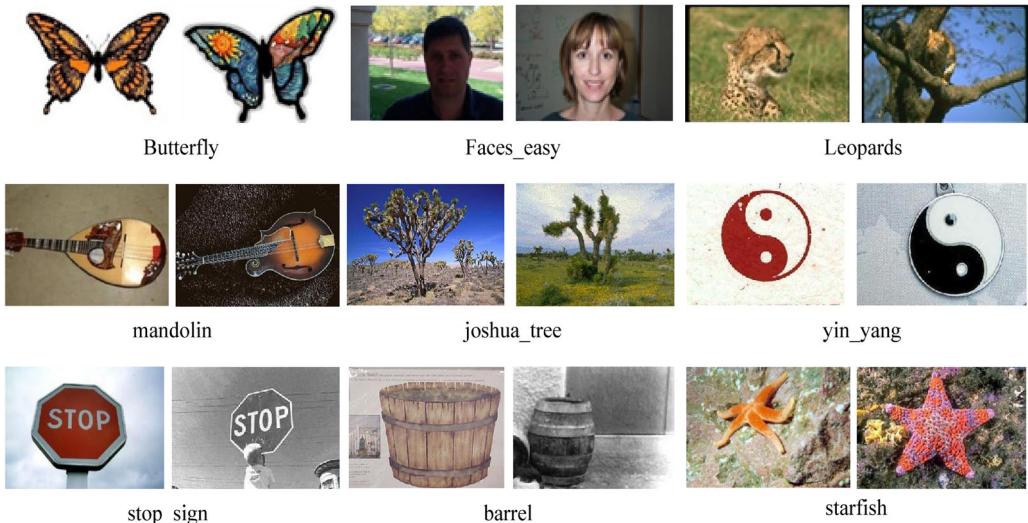


Fig. 3 Sample images for the Scene15 dataset

Oxford Flowers dataset is composed of 17 flower categories with 80 images per class. The dataset contains totally 1360 images in which intra-class variation is sometimes greater than that of the inter-class between two flower species. With reference to corresponding experiments using Oxford Flowers dataset, 40 images per category are employed for training and the rest for testing. As illustrated in Fig. 4, it can be seen as a challenging dataset for classification task from sample images.

The Caltech101 dataset contains 9144 images belonging to 101 object classes and an additional category of background images with high intra-class appearance and shape variability. The number of images in each class varies between 31 and 800. Following the common experiment setup for Caltech101, we train on 30 images and evaluate the performance using the rest for testing. Some representative sample images are shown in Fig. 5.

The proposed approach is compared with more than ten kinds of state-of-the-art methods in terms of classification accuracy in this part. Besides, the processor of our PC is Intel Core i5 and frequency is 3.1 GHZ. In order to cope with the high-dimensional p.d.f features, the configuration of RAM

**Fig. 4** Sample images for the Oxford Flowers dataset**Fig. 5** Sample images for the Caltech101 dataset

is 32 GB and the operating system is 64-bit Windows 7. All codes are written in Matlab.

5.1 Evaluations with respect to three types of kernel functions in the first layer on diverse datasets

For the sake of achieving good performance in kernel methods, we conduct the first study on three commonly used kernel functions. It is very important to find appropriate kernels associated with specific image features. As stated in Sect. 3, we choose polynomial, GRBF and HIK functions to evaluate the performance with two different features includ-

ing SPM and p.d.f gradient-based orientation histogram (p.d.f feature) in the first layer, respectively. According to the typical parameters selection used in [48, 54], we choose $a = 0.8$, $b = 3$, $d = 1.3$ and vary the number of atoms in this experiment. In addition, each result is averaged out by five times with the sparsity level $T = 120$ for training and testing KSR. Figure 6 summarizes the classification accuracy with a varied number of dictionary atoms when different kernel functions are exploited on SPM features.

It is easy to observe that HIK significantly outperform the other two candidates when SPM features are

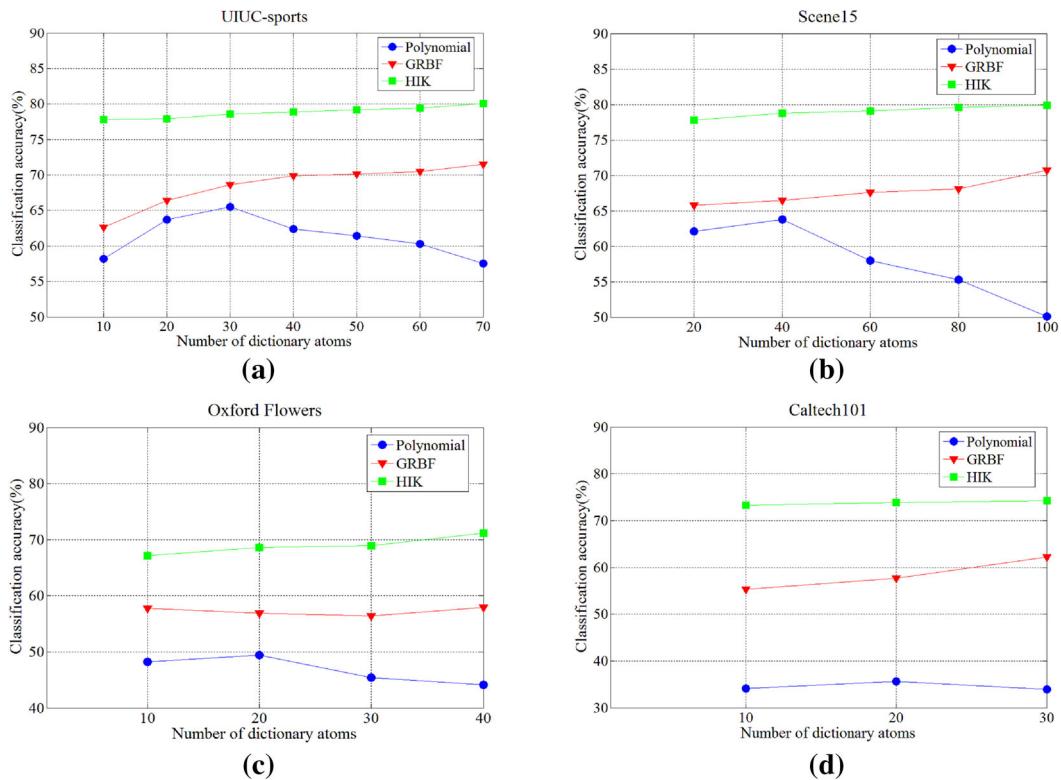


Fig. 6 Comparison of classification accuracy with three different kernel functions using SPM features on four datasets (UIUC-Sports, Scene15, Oxford Flowers and Caltech101)

extracted for KSR on four datasets. The classification accuracy rises consistently as the number of dictionary atoms increases on all datasets while the performance of polynomial and GRBF kernels is unstable with regard to the SPM features. Meanwhile, it should be noted that the comparison results qualitatively show tendency of kernel selection and it does not necessarily mean that the results are the optimal value given a fixed number of atoms in each dictionary.

Similarly, in the next stage, we further study the performance of different kernel functions on p.d.f features in the first layer. In this experiment, the parameters selection of kernels remain the same as described above and the results are obtained by 5 times repeated test, while the sparsity level T is changed to be 180 for training and testing KSR, respectively. With the varying number of dictionary atoms, the comparison of classification accuracy on four datasets can be seen from Fig. 7.

Compared with the results using SPM features, the performance of GRBF kernel on p.d.f features is obviously superior to that of HIK and polynomial kernel as increasing the number of dictionary atoms. More importantly, the experimental results provided by the first layer is just slightly inferior to that of the recent work [44] which utilizes SMO for SVM on UIUC-Sports and Scene15 datasets.

5.2 Evaluations with respect to the sparsity level of KSR in the first layer on diverse datasets

According to the analysis of the results in part A, it suggests that HIK and GRBF kernel have an edge over other kernel functions on SPM and p.d.f features, respectively. For this reason, we furthermore attempt to delve deeper toward understanding other factors affecting the classification accuracy in the first layer. Given that HIK has no additional parameters for KSR, we only need to adjust sparsity level T to evaluate the performance on four datasets while the dictionary size keeps the maximum number in this step. Figure 8 displays the classification accuracy for different sparsity levels.

As shown in Fig. 8, the sparsity level ranges from 60 to 260 and the classification accuracy peaks when T is set to be 160, 240, 120 and 200 on UIUC-Sports, Scene15, Oxford Flowers and Caltech101, respectively. While the sparsity level increases further, the accuracy goes down in different degrees. Therefore, the optimal setting of sparsity level for KSR can be used in feature concatenation.

In contrast to the HIK on SPM features, the performance of GRBF kernel on p.d.f features is influenced by both the mean of the pairwise distance d and the number of nonzero elements T . Hence, we alter one parameter while the other one remains unchanged in this experiment. Figure 9a shows

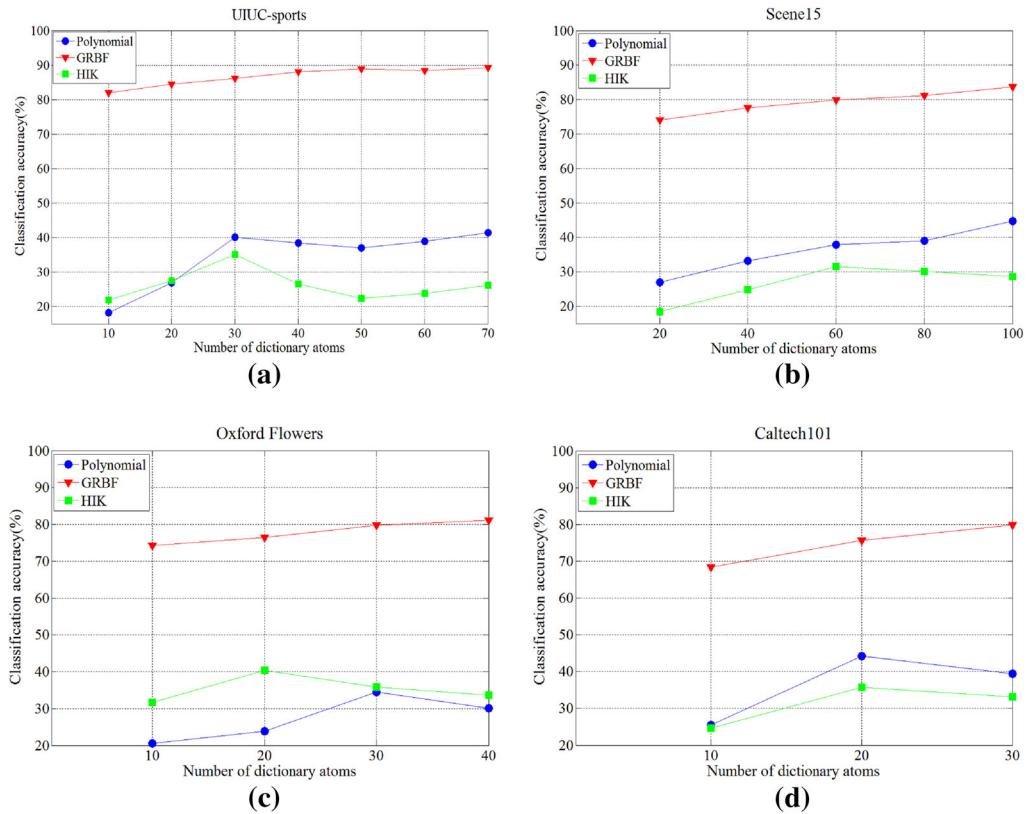


Fig. 7 Comparison of classification accuracy with three different kernel functions using p.d.f features on four datasets (UIUC-Sports, Scene15, Oxford Flowers and Caltech101)

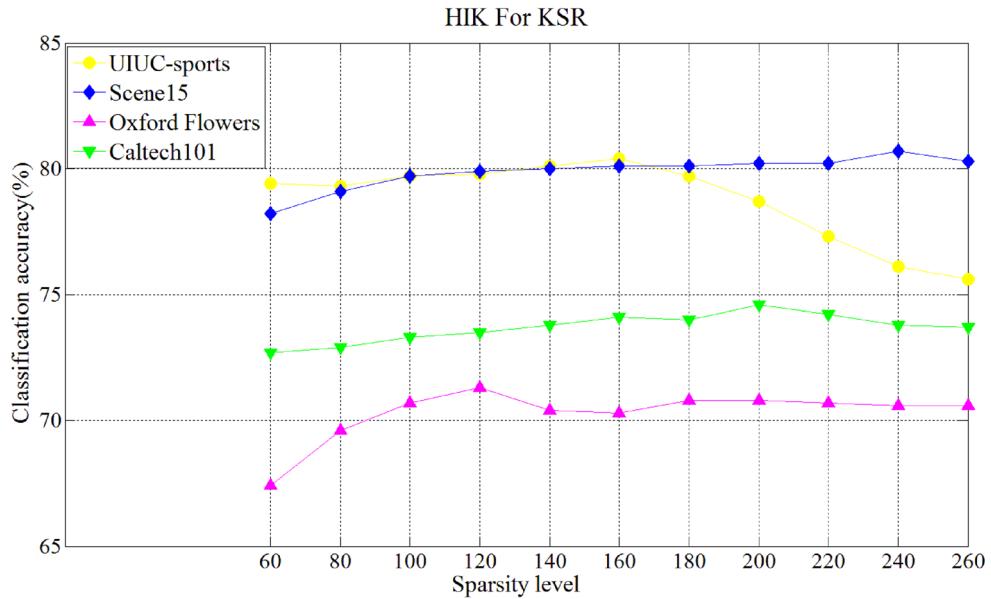
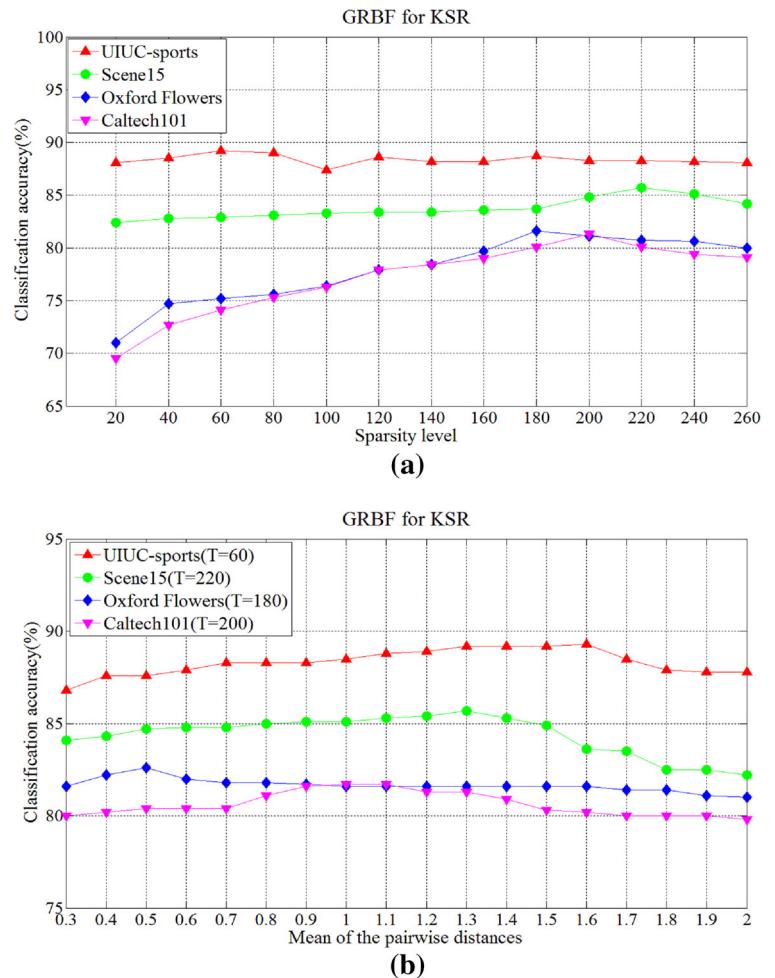


Fig. 8 Comparison of classification accuracy for different sparsity levels of KSR on four datasets (UIUC-Sports, Scene15, Oxford Flowers and Caltech101)

the comparison of classification accuracy on diverse datasets when d is fixed to be 1.3 and the number of dictionary atoms remains the same as the last step. We observe that the accura-

cies increase steadily and reach a maximum at $T = 220, 180, 200$ on Scene15, Oxford Flowers and Caltech101 datasets, respectively. From these points onwards, they decline grad-

Fig. 9 **a** Comparison of classification accuracy for different sparsity levels ($d = 1.3$). **b** Comparison of classification accuracy for different mean of the pairwise distances



ually for the rest of sparsity levels, whereas the accuracy fluctuates from 20 to 260 and reach the peak at $T = 60$ on UIUC-Sports dataset. Thus, we choose those sparsity levels indicating the best performance for the next step.

Empirically, the parameter selection of d ranges from 0.3 to 2.0 in the steps of 0.1. Figure 9b demonstrates that the classification accuracies reach a maximum at $d = 1.6, 1.3, 0.5, 1.0$ on four datasets, respectively. It should be noted that the performance on Caltech101 of p.d.f features using GRBF kernel for KSR is marginally better than that of C-kernel KSVD which combines 39 different features using the boosting method.

5.3 Evaluations of hierarchical feature concatenation-based KSR on diverse datasets

Similar to the evaluations of the two parts above in Sect. 4, the KKSVD and BKOMP in the second layer is built on top of the concatenated outputs of the KSR on SPM and p.d.f features in the first layer. To begin with, GRBF kernel function is employed for this experiment. According to the analysis of the experimental results in the first layer, the dictionary size

remains the maximum for each dataset which provides the best performance in terms of classification accuracy. Next, we select $d = 2.0$ and change T ranging from 20 to 280 to balance the sparsity and efficiency. The relationship between T and classification accuracy is shown in Fig. 10.

We see that our approach achieves promising results in a wide range of sparsity levels and $T = 160, 240, 240, 260$ can be chosen for the next stage of experiment. The range of parameter d is extended from 0.5 to 5.0 in the steps of 0.5 and the classification accuracies on four datasets are illustrated in Fig. 11.

As can be seen from Fig. 11, compared with the pattern using fixed d , adjustment of this parameter also has a large impact on the performance of GRBF kernel for concatenated features. In contrast to one-layer KSR for categorization tasks, the proposed method substantially improves classification accuracy employing appropriate kernel functions and parameters. Specifically, our approach outperforms HIK on SPM features by 14.2, 12.8, 18.9, 14.6 % and GRBF kernel on p.d.f features by 5.3, 7.8, 7.6, 7.5 %, respectively, on four commonly used benchmarks. The comparison of the performance can be directly visualized in Fig. 12.

Fig. 10 Comparison of classification accuracy for different sparsity levels ($d = 2.0$)

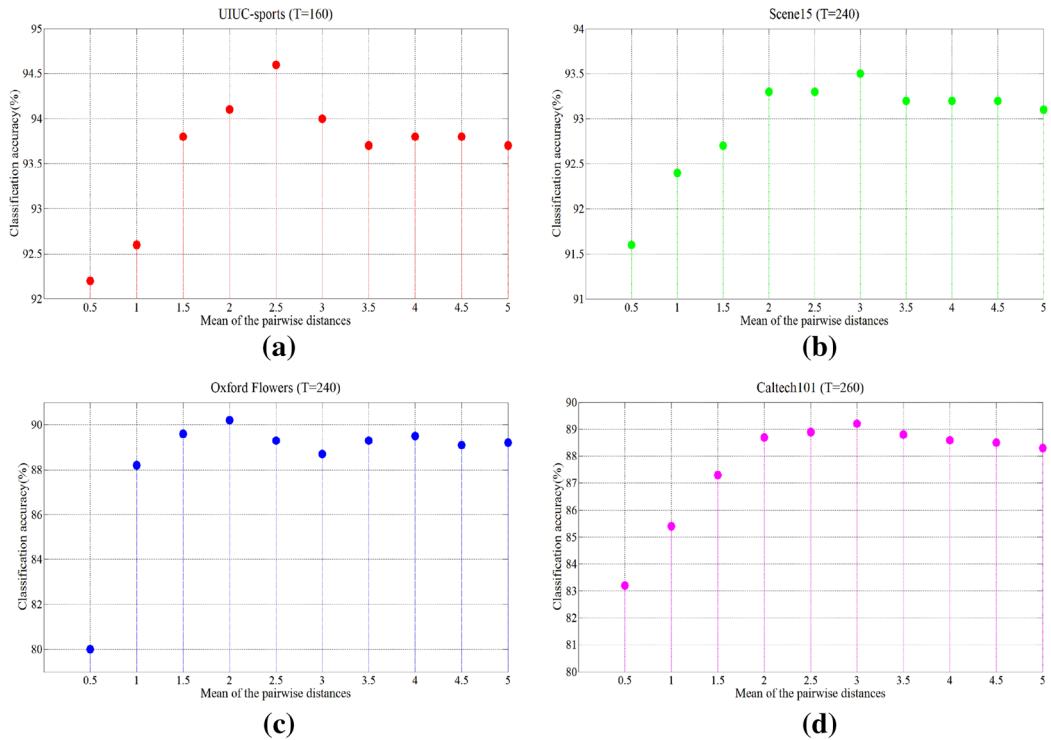
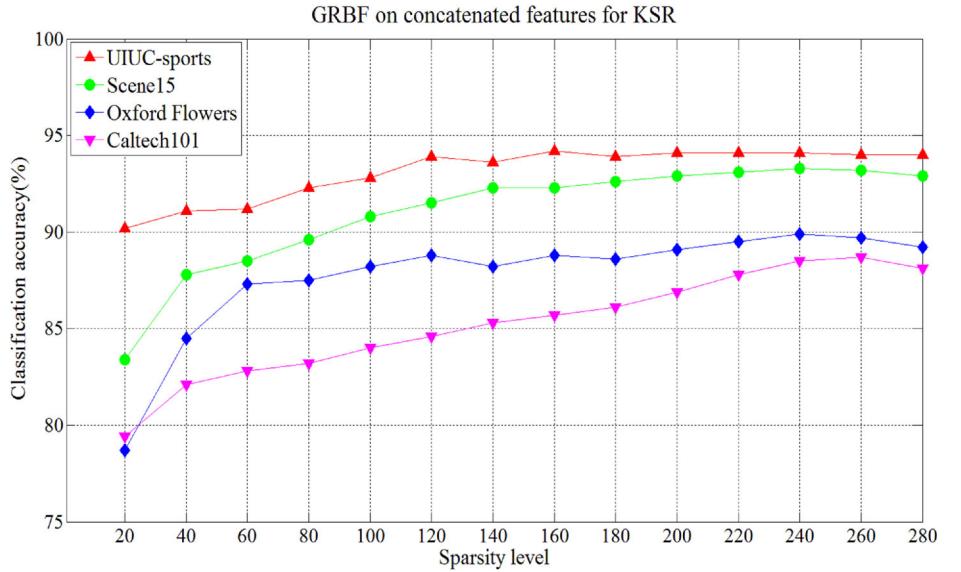


Fig. 11 Comparison of classification accuracy for different mean of the pairwise distances using GRBF on concatenated features (UIUC-Sports, Scene15, Oxford Flowers and Caltech101)

Finally, we compare our proposed scheme with MKL methods employing multiple image features and other state-of-the-art approaches on UIUC-Sports, Scene15, Oxford Flowers and Caltech101, respectively, in Table 3, 4, 5, 6.

The experimental results are listed in the Table 3. We can see that our proposed method performs much better than K^{CL} [55] and KSRSPM-HIK [18]. The former combines

the advantage of the high-level representation of the Object Bank with kernel methods. While the latter has shown success of finding sparse representations of nonlinear features with HIK, it solves nonlinear regression with l_1 -norm. In addition, our scheme performs slightly better than SKDES [56] employing a useful framework to embed the image label into the design of patch-level kernel descriptors.

Fig. 12 Comparison of the performance between hierarchical and one-layer KSR for classification

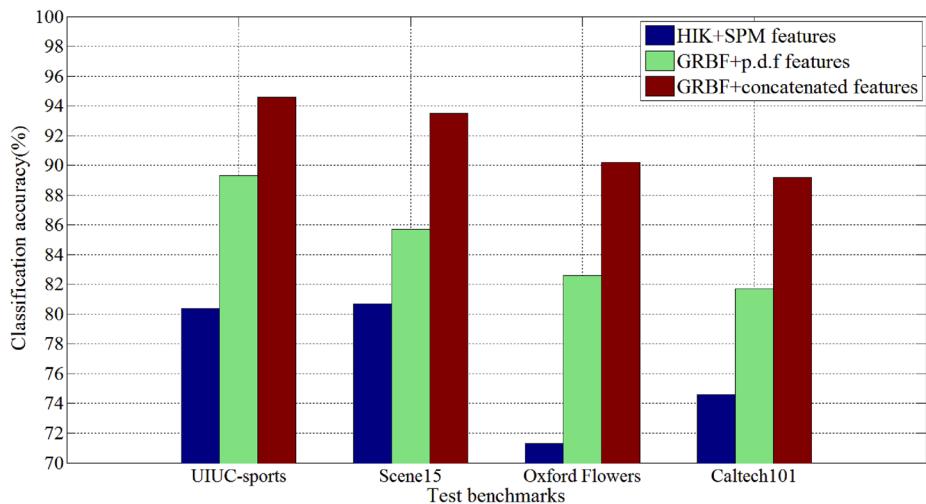


Table 3 Comparison of the classification accuracies on UIUC-Sports dataset

| Method | Classification accuracy (%) |
|-----------------|-----------------------------|
| K^{CL} [55] | 86.02 |
| KSRSPM-HIK [18] | 86.85 |
| SKDES [56] | 91.0 |
| Proposed | 94.6 |

Table 4 Comparison of the classification accuracies on Scene15 dataset

| Method | Classification accuracy (%) |
|---------------|-----------------------------|
| GS-MKL [30] | 86.5 |
| PCK-LMML [57] | 87.3 |
| SKDES [56] | 88.7 |
| K^{CL} [55] | 88.81 |
| S-LMML [58] | 89.13 |
| GA-MKL [59] | 89.32 |
| Proposed | 93.5 |

As for the results of Scene15 benchmark from the Table 4, the proposed pattern outperforms recent MKL methods including GS-MKL [30], PCK-LMML [57], S-LMML [58] and GA-MKL [59] by about 7.0, 6.2, 4.4 and 4.2 %. Besides, the classification accuracies of the other two approaches described above are marginally below our scheme by 4.8 and 4.7 %.

Table 5 compares the performance of our presented method to some other methods which have shown impressive results on Oxford Flowers dataset. Our scheme achieves accuracy of 90.2 % which is slightly better than that of LP- β [15], MKSR [60], MKC [61], and KMTJSRC-CG [62]. It is worthy to note that all the competitors adopt relatively com-

Table 5 Comparison of the classification accuracies on Oxford Flowers dataset

| Method | Classification accuracy (%) |
|------------------|-----------------------------|
| LP- β [15] | 85.5 |
| MKSR [60] | 86.3 |
| MKC [61] | 88.33 |
| KMTJSRC-CG [62] | 88.9 |
| Proposed | 90.2 |

Table 6 Comparison of the classification accuracies on Caltech101 dataset

| Method | Classification accuracy (%) |
|------------------|-----------------------------|
| K^A [55] | 65.5 |
| MKL-SRC [54] | 75.7 |
| PCK-LMML [57] | 77.8 |
| SKDES [56] | 79.2 |
| C-KKSVD [17] | 80.1 |
| LP- β [15] | 82.1 |
| MKSR [60] | 82.9 |
| GA-MKL [59] | 83.12 |
| GS-MKL [30] | 84.6 |
| Proposed | 89.2 |

plex multiple discriminative feature combination to boost performance.

Table 6 lists the comparison results over ten kinds of algorithms, in which MKL-SRC [54], PCK-LMML [57], LP- β [15], MKSR [60], GA-MKL [59], GS-MKL [30] use MKL technique for classification. As to Caltech101 benchmark, 89.2 % achieved by our method is far above K^A [55], SKDES [56] and C-KKSVD [17] by 23.7, 10.0 and 9.1 %, respectively. It should be noted that [17] combines 39 image features

using boosting pattern and also employs KKSVD for dictionary learning.

6 Conclusions

In this paper, we have presented hierarchical feature concatenation-based KSR for image categorization. The proposed method using hierarchical concatenation of kernelized p.d.f gradient-based orientation histogram and SPM features outperforms the approach employing each kernelized feature individually. The experimental results demonstrate that the GRBF kernel function and the parameter-free HIK function are appropriate for p.d.f and SPM features, respectively, in our HKOMP scheme. In the second layer, the compact concatenated KSR with GRBF kernel function can learn more discriminative sparse codes for categorization task which is shown on classification accuracy compared with some state-of-the-art MKL algorithms and other competitive kernel learning based methods. More importantly, the results prove the effectiveness of implementation on more powerful rather than multiple image features for kernel learning based method.

Acknowledgments This project is supported by the National Program on Key Basic Research Project (No. 2014CB340403) and Natural Science Foundation of Tianjin (No. 15JCYBJC15500). The authors would like to be grateful to the editors' and reviewers' valuable comments which improved the quality of this paper.

References

- Zhang, L., Zhao, Y., Zhu, Z.: Extracting shared subspace incrementally for multi-label image classification. *Vis. Comput.* **30**(12), 1359–1371 (2014)
- Liu, X., Shi, Z., Shi, Z.: A co-boost framework for learning object categories from Google Images with 1st and 2nd order features. *Vis. Comput.* **30**(1), 5–17 (2013)
- Ji, R., Gao, Y., Hong, R., Liu, Q., Tao, D., Li, X.: Spectral-spatial constraint hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **52**(3), 1811–1824 (2014)
- Gao, Y., Wang, M., Zha, Z., Shen, J., Li, X., Wu, X.: Visual-textual joint relevance learning for tag-based social image search. *IEEE Trans. Image Process.* **22**(1), 363–376 (2013)
- Gao, Y., Wang, M., Tao, D., Ji, R., Dai, Q.: 3D object retrieval and recognition with hypergraph analysis. *IEEE Trans. Image Process.* **21**(9), 4290–4303 (2012)
- Gao, Y., Wang, M., Ji, R., Wu, X., Dai, Q.: 3D object retrieval with hausdorff distance learning. *IEEE Trans. Ind. Electron.* **61**(4), 2088–2098 (2014)
- Guan, T., He, Y., Duan, L.: Efficient BOF generation and compression for on-device mobile visual location recognition. *IEEE Multimed.* **21**(2), 32–41 (2014)
- Guan, T., He, Y., Gao, J., Yang, J., Yu, J.: On-device mobile visual location recognition by integrating vision and inertial sensors. *IEEE Trans. Multimed.* **21**(2), 32–41 (2014)
- Guan, T., Wang, Y., Duan, L., Ji, R.: On-device mobile landmark recognition using binarized descriptor with multifeature fusion. *ACM Trans. Intell. Syst. Technol.* **7**(1), 1–28 (2015)
- Zhao, Y., Yang, J.: Hyperspectral image denoising via sparse representation and low-rank constraint. *IEEE Trans. Geosci. Remote Sens.* **53**(1), 296–308 (2015)
- Zhao, Z., Glotin, H., Xie, Z., Gao, J., Wu, X.: Cooperative sparse representation in two opposite directions for semi-supervised image annotation. *IEEE Trans. Image Process.* **21**(9), 4218–4231 (2012)
- Chiang, C.-K., Liu, C.-H., Duan, C.-H., Lai, S.-H.: Learning component-level sparse representation for image and video categorization. *IEEE Trans. Image Process.* **22**(12), 4775–4787 (2013)
- Wang, L., Yan, H., Lv, K., Pan, C.: Visual tracking via kernel sparse representation with multikernel fusion. *IEEE Trans. Circuits Syst. Video Technol.* **24**(7), 1132–1141 (2014)
- Zhang, L., Zhou, W., Chang, P.-C., Liu, J., Yan, Z., Wang, T., Li, F.: Kernel sparse representation-based classifier. *IEEE Trans. Signal Process.* **60**(4), 1684–1695 (2012)
- Gehler, P., Nowozin, S.: On feature combination for multiclass object classification. In: IEEE International Conference on Computer Vision, pp. 221–228 (2009)
- Zheng, J., Huang, Q., Chen, S., Wang, W.: Efficient kernel discriminative common vectors for classification. *Vis. Comput.* **31**(5), 643–655 (2015)
- Nguyen, H., Patel, V., Nasrabadi, N., Chellappa, R.: Design of non-linear kernel dictionaries for object recognition. *IEEE Trans. Image Process.* **22**(12), 5123–5135 (2013)
- Gao, S., Tsang, I.W., Chia, L.-T.: Sparse representation with kernels. *IEEE Trans. Image Process.* **22**(2), 423–434 (2013)
- Jian, M., Jung, C.: Class-discriminative kernel sparse representation-based classification using multi-objective optimization. *IEEE Trans. Signal Process.* **61**(18), 4416–4427 (2013)
- Shawe-Taylor, J., Cristianini, N.: *Kernel Methods for Pattern Analysis*. Cambridge University Press, Cambridge (2004)
- Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM International Conference on Image Video Retrieval, pp. 401–408 (2007)
- Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
- Tuytelaars, T.: Dense interest points. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2281–2288 (2010)
- Boureau, Y., Bach, F., Yann, L., Ponce, J.: Learning mid-level features for recognition. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2559–2566 (2010)
- Sonnenburg, S., Rätsch, G., Schäfer, C., Schölkopf, B.: Large scale multiple kernel learning. *J. Mach. Learn. Res.* **7**, 1531–1565 (2006)
- Rakotomamonjy, A., Bach, F., Canu, S., Grandvalet, Y.: More efficiency in multiple kernel learning. In: Proceedings of the 24th International Conference on Machine Learning, pp. 775–782 (2007)
- Vedaldi, A., Gulshan, V., Varma, M., Zisserman, A.: Multiple kernels for object detection. In: IEEE International Conference on Computer Vision, pp. 606–613 (2009)
- Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., Torralba, A.: SUN database: large-scale scene recognition from abbey to zoo. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 3485–3492 (2010)
- Patterson, G., Hays, J.: SUN attribute database: discovering, annotating, and recognizing scene attributes. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2751–2758 (2012)
- Yang, J., Tian, Y., Duan, L.-Y., Huang, T., Gao, W.: Group-sensitive multiple kernel learning for object recognition. *IEEE Trans. Image Process.* **21**(5), 2838–2852 (2012)

31. Jain, A., Vishwanathan, S.V.N., Varma, M.: SPF-GMML: generalized multiple kernel learning with a million kernels. In: Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining, pp. 750–758 (2012)
32. Gönen, M., Alpaydin, E.: Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **12**, 2211–2268 (2011)
33. Ojala, T., Pietikainen, M., Maenpää, T.: Multiresolution grayscale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002)
34. Li, C., Zhou, W., Yuan, S.: Iris recognition based on a novel variation of local binary pattern. *Vis. Comput.* **31**(10), 1419–1429 (2015)
35. Tuzel, O., Porikli, F., Meer, P.: Human detection via classification on riemannian manifolds. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
36. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
37. Murtza, I., Abdullah, D., Khan, A., Arif, M., Mirza, S.M.: Cortex-inspired multilayer hierarchy based object detection system using PHOG descriptors and ensemble classification. *Vis. Comput.* 1–14 (2015)
38. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* **42**(3), 145–175 (2001)
39. Pinto, N., Cox, D.D., Dicarlo, J.J.: Why is real-world visual object recognition hard. *PLOS Comput. Biol.* **4**(1), e27 (2008)
40. Berg, A., Malik, J.: Geometric blur for template matching. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 607–614 (2001)
41. Varma, M., Ray, D.: Learning the discriminative power-invariance trade-off. In: IEEE International Conference on Computer Vision, pp. 1–8 (2007)
42. Bucak, S., Jin, R., Jain, A.: Multiple kernel learning for visual object recognition: a review. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(7), 1354–1369 (2014)
43. Aiolfi, F., Donini, M.: EasyMKL: a scalable multiple kernel learning algorithm. *Neural Comput.* **16**9, 215–224 (2015)
44. Kobayashi, T.: BFO meets HOG: feature extraction based on histograms of oriented p.d.f. gradients for image classification. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 947–954 (2013)
45. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2169–2178 (2006)
46. Yin, J., Liu, Z., Jin, Z., Yang, W.: Kernel sparse representation based classification. *Neural Comput.* **7**7, 120–128 (2011)
47. Li, H., Gao, Y., Sun, J.: Fast kernel sparse representation. In: Proceedings of International Conference on Digital Image Computing Techniques and Applications, pp. 72–77 (2011)
48. Nguyen, H., Patel, V., Nasrabadi, N.M., Chellappa, R.: Kernel dictionary learning. In: Proceedings of IEEE International Conference on Acoustics, Speech, Signal Process, pp. 2021–2024 (2012)
49. Yu, K., Lin, Y., Lafferty, J.: Learning image representations from the pixel level via hierarchical Sparse coding. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1713–1720 (2011)
50. Bo, L., Ren, X., Fox, D.: Hierarchical matching pursuit for image classification: architecture and fast algorithms. In: Advances in neural information processing systems, pp. 2115–2123
51. Bo, L., Ren, X., Fox, D.: Multipath sparse coding using hierarchical matching pursuit. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 660–667 (2013)
52. Liu, B., Liu, J., Bai, X., Lu, H.: Regularized hierarchical feature learning with non-negative sparsity and selectivity for image classification. In: Proceedings of the 22nd International Conference on Pattern Recognition, pp. 4293–4298 (2014)
53. Wu, J., Rehg, J.M.: Beyond the Euclidean distance: creating effective visual codebooks using the histogram intersection kernel. In: IEEE International Conference on Computer Vision, pp. 630–637 (2009)
54. Shrivastava, A., Patel, V., Chellappa, R.: Multiple kernel learning for sparse representation-based classification. *IEEE Trans. Image Process.* **23**(7), 3013–3024 (2014)
55. Zhang, L., Zhen, X., Shao, L.: Learning object-to-class kernels for scene classification. *IEEE Trans. Image Process.* **23**(8), 3241–3253 (2014)
56. Wang, P., Wang, J., Zeng, G., Xu, W., Zha, H., Li, S.: Supervised kernel descriptors for visual recognition. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2858–2865 (2013)
57. Han, Y., Liu, G.: Probability-confidence-kernel-based localized multiple kernel learning with l_p norm. *IEEE Trans. Syst. Man Cybern. B* **42**(3), 827–837 (2012)
58. Han, Y., Yang, K., Ma, Y., Liu, G.: Localized multiple kernel learning via sample-wise alternating optimization. *IEEE Trans. Cybern.* **44**(1), 137–148 (2014)
59. Yan, S., Xu, X., Xu, D., Lin, S., Li, X.: Image classification with densely sampled image windows and generalized adaptive multiple kernel learning. *IEEE Trans. Cybern.* **45**(3), 395–404 (2015)
60. Thiagarajan, J., Ramamurthy, K., Spanias, A.: Multiple kernel sparse representations for supervised and unsupervised learning. *IEEE Trans. Image Process.* **23**(7), 2905–2915 (2014)
61. Nilsback, M.-E., Zisserman, A.: Automated flower classification over a large number of classes. In: Proceedings of the 6th Indian Conference on Computer Vision, Graphics and Image Processing, pp. 722–729 (2008)
62. Yuan, X.-T., Yan, S.: Visual classification with multi-task joint sparse representation. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 3493–3500 (2010)

Bo Wang is pursuing the Ph.D. degree in the School of Electronic Information Engineering, Tianjin University, Tianjin, China. His research interests include image processing, pattern recognition, computer vision and machine learning.





Jichang Guo received the B.S. degree in Department of Electronic Engineering from Nanjing University of Science and Technology, Nanjing, China, in 1988, the MSc degree and the Ph.D. degree in School of Electronic Information Engineering from Tianjin University, Tianjin, China, in 1993 and 2003, respectively. He is currently a Professor in School of Electronic Information Engineering at Tianjin University. His research area includes digital image processing, filter design and intelligent information processing, in which he has published about 70 scientific papers.



Chongyi Li pursuing the Ph.D. degree in the School of Electronic Information Engineering, Tianjin University, Tianjin, China. His current research focuses on digital image processing, computer vision, and especially on image enhancement.



Yan Zhang was born in 1982 in Shandong Province, China. She received the M.S. degree in circuit and system from the school of electronic information engineering, Tianjin University, Tianjin, China, in 2007. Currently, she is a Ph.D. candidate in signal and information processing at the Tianjin University. And she is a lecturer in Tianjin Chengjian University. Her research interests include digital image processing and intelligent information processing.