

RESEARCH

Open Access



A feature fusion based localized multiple kernel learning system for real world image classification

Fatemeh Zamani*  and Mansour Jamzad

Abstract

Real-world image classification, which aims to determine the semantic class of un-labeled images, is a challenging task. In this paper, we focus on two challenges of image classification and propose a method to address both of them simultaneously. The first challenge is that representing images by heterogeneous features, such as color, shape and texture, helps to provide better classification accuracy. The second challenge comes from dissimilarities in the visual appearance of images from the same class (intra class variance) and similarities between images from different classes (inter class relationship). In addition to these two challenges, we should note that the feature space of real-world images is highly complex so they cannot be linearly classified. The kernel trick is efficacious to classify them. This paper proposes a feature fusion based multiple kernel learning (MKL) model for image classification. By using multiple kernels extracted from multiple features, we address the first challenge. To provide a solution for the second challenge, we use the idea of a localized MKL by assigning separate local weights to each kernel. We employed spatial pyramid match (SPM) representation of images and computed kernel weights based on X^2 kernel. Experimental results demonstrate that our proposed model has achieved promising results.

Keywords: Image classification, Spatial pyramid matching, Localized multiple kernel learning, Kernel local weighting, Feature fusion

1 Introduction

The complex structure of human visual system and the heavy processes performed in the brain when looking at an image provide impressive ability to recognize real images in a fraction of a second. Although real world image classification, which is the focus of this paper, seems to be trivial for humans, it is a challenging task in computer vision. In recent years, image classification has attracted a lot of attention in computer vision due to the rapid improvement of intelligent robots and the need for processing images.

There is a very rich literature on image classification including methods based on bag of word [1, 2], Sparse representation [3–7], and Deep learning [8–10]. We should point out that nonlinear classifiers, including kernel based ones, have gained more attention due to their high performance compared to linear classifiers [5, 7, 9].

Classifying real world images is a challenging task. Following are the two challenges which this paper concentrates

on. First, images cannot be described precisely by one single feature; therefore, they should be represented by multiple features such as color, shape and texture. Second, the intra class variance (dissimilarity between images in the same class) and inter class relationship (similarities between images from different classes) are large. The mentioned challenges are discussed in the following sub-sections.

1.1 The effectiveness of using multiple features

Images are informative in different aspects like color, shape and texture. Describing images with multiple features rather than a single feature, results in a more accurate classifier. For example, an approach is proposed in [11] which describes an image by means of multiple bag of word features and designs a classifier based on them. Also, some kernel based classifiers are proposed based on multiple features [12–16].

1.2 Large intra class variance and inter class relationship

The second principal challenge in real world image classification is the existence of large intra class variance

* Correspondence: f_zamani@ce.sharif.edu

Computer Engineering Department, Sharif University of Technology, Tehran, Iran

and large interclass relationship between images. Even if we use multiple features, there are images in a class which could be considered dissimilar (large intra class variance). Moreover, there are images from different classes that may be classified to one class (large inter class relationship). Fig. 1 is an illustration of the second challenge.

In addition to the two described challenges, we should note that feature spaces of real world images are complex, so they cannot be linearly classified. Kernel based methods have achieved major success in building nonlinear classifiers [17]. A multiple kernel learning (MKL) framework proposed by Lanckriet et al. is considered as one of the most powerful classifiers [18]. To classify data, MKL considers a linearly weighted sum of kernels instead of a single kernel. By using MKL we can combine different kernels. Each kernel is computed based on an individual feature (for example, a color based kernel describes the color information of an image). In this way, the first challenge is addressed.

In the standard framework of MKL, as stated above, the computed weights of kernels are the same for all samples. This means that each kernel has a fixed share in deciding the class of each test image. With respect to the second challenge, a more accurate classifier will be achieved if the share of each kernel is not similar; and its weight is computed based on its efficiency in classification of samples. For example, in the first row of Fig. 1, to prevent misclassification the weight of the color based

kernel should be reduced while the weights of other kernels should be increased.

Gonen et al. proposed a localized multiple kernel learning (LMKL) framework which computes non-uniform weights for kernels based on their location in the feature space [19]. LMKL is briefly reviewed in section 2. To address both challenges mentioned in subsections 1.1 and 1.2, we propose a feature fusion version of the original LMKL. A comparison between a single kernel based on SVM, MKL, LMKL, and the feature fusion based LMKL is illustrated in Fig. 2. The block diagram of our proposed system is depicted in Fig. 3. Our experiments on Caltech 101 and Caltech 256 achieved promising results.

The rest of paper is organized as follows. A brief review about LMKL is given in section 2. In section 3 the proposed algorithm is discussed in detail. The experimental results are given and analyzed in section 4. Finally, we conclude the paper in section 5.

2 LMKL related work

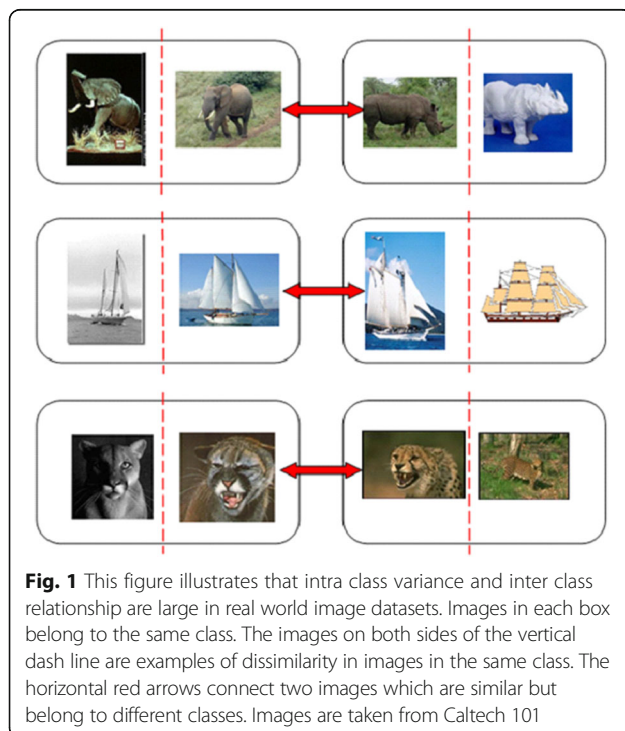
In this section, we give a brief review about the related work of localized multiple kernel learning (LMKL) which is an extension of the MKL framework. The original MKL computes fixed weight for each kernel by embedding kernel weights in the SVM optimization problem and then constructs a single kernel by summing up the weighted kernels [18]. In [12], fixed weights for kernels are computed by a slight modification of MKL framework. It extracts heterogeneous features from data then a group of kernels is assigned to each feature. By using a group lasso regularization method, only a few kernels are selected for each feature.

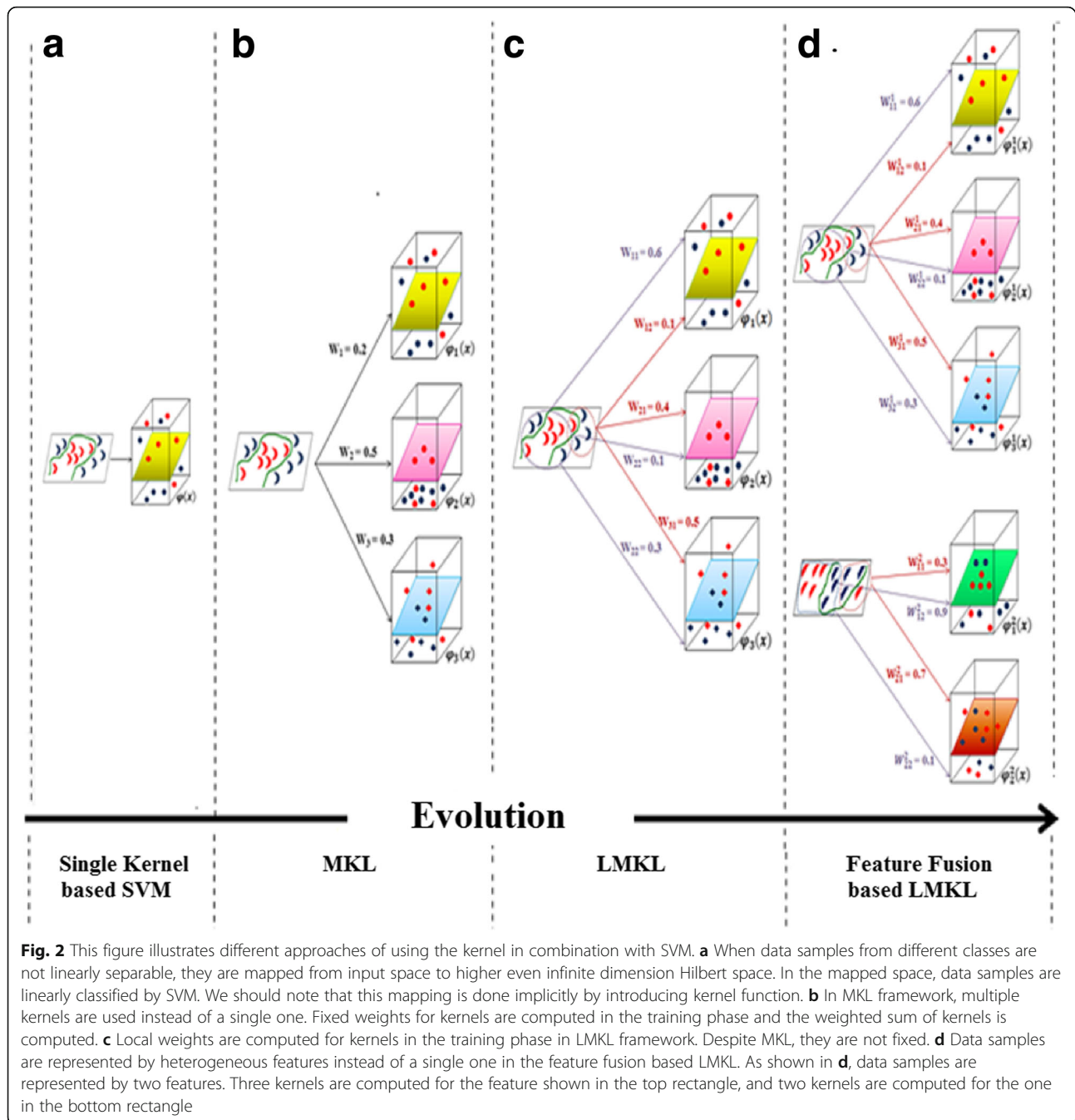
Some other works dedicate fixed weights to kernels without using the standard MKL framework. Gu et al. computed fixed weights for kernels by projecting them in the maximum variance direction [20]. Wang et al. computed optimal fixed kernel weights by finding the best projective direction which results in maximum separation between kernels in RKHS (Reproducing Kernel Hilbert Space) [21].

There are some approaches which combine kernels in a nonlinear manner while the weight of each kernel is fixed. For example, in [22], all weighted kernel matrices are combined by Hadamard product while the kernel matrix and its corresponding weight are powered by an identical number. Algorithms which combine weighted kernels are reviewed and discussed in [23].

As discussed in section 1.2, in problems like image classification, it is more beneficial to use variable weights for each kernel. Some algorithms which compute variable weights for kernels are discussed below.

Lewis et al. combined kernels in a non-stationary manner in a framework of maximum entropy discrimination [24]. Lee et al. proposed a method to combine kernels without





learning distinct weights for kernels [25]. In this method, the local impact of each kernel is directly considered in the process of margin maximization. Gönen et al. designed a nonlinear framework which computes separate kernel weights for each data point based on nonlinear gating functions [19]. Yang et al. defined interclass clusters of samples and found the optimal kernel combinations for each cluster in an image classification task [26]. In [19, 26] the authors suggested to partition the space linearly.

Kannao et al. allowed nonlinear boundary between clusters of the space [27]. They computed a linear kernel weight per cluster in a pre-process step without considering the sample labels.

Despite the functionality of computing variable weights for kernels, few works in image classification are based on this approach. Lu et al. proposed a Localized Multiple Kernel Metric Learning approach to classify images taken from varying viewpoints or under varying illuminations

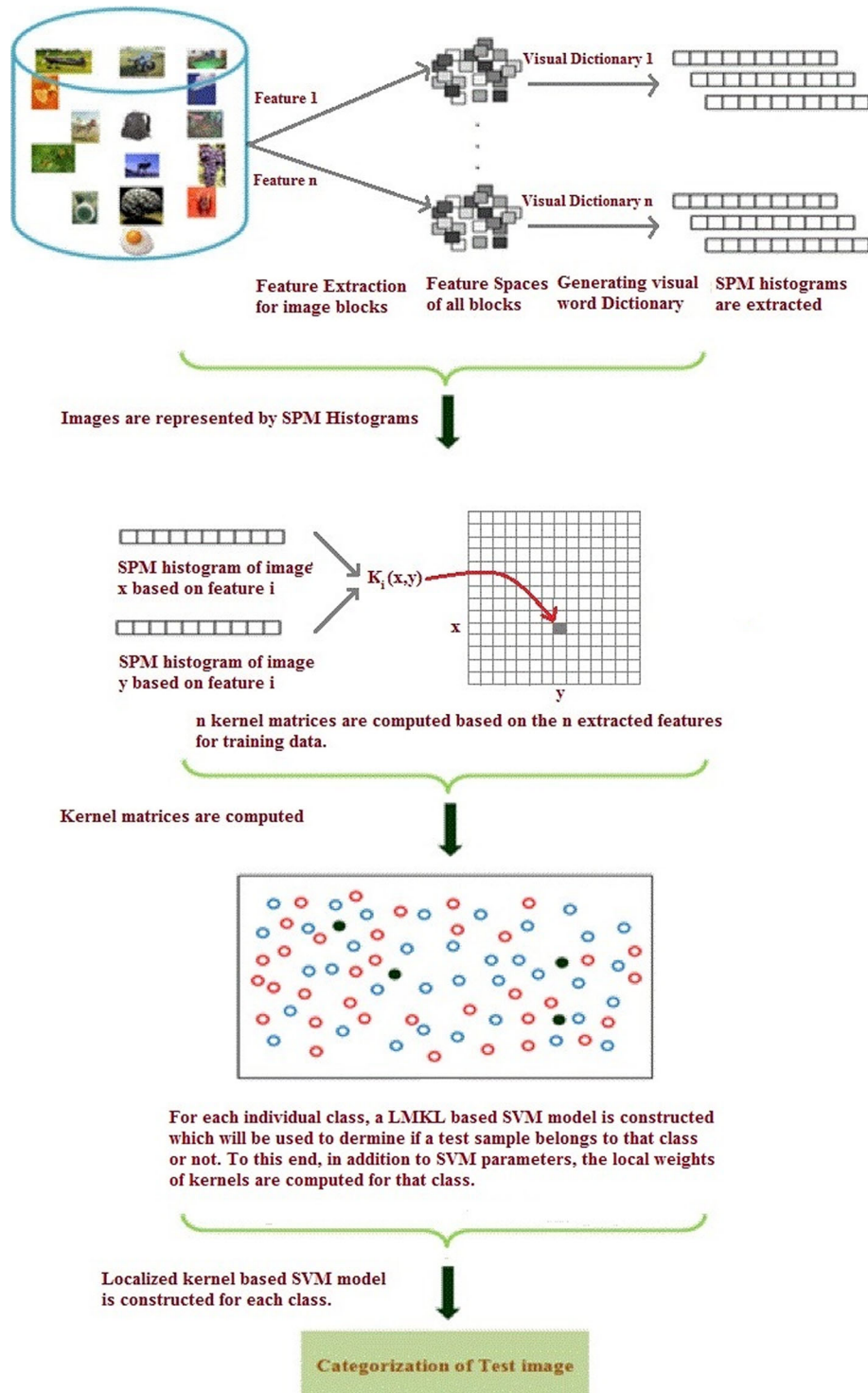


Fig. 3 Block diagram of the proposed feature fusion-based LMKL

[28]. Fan et al. considered the relationship between global and local structures of features [29]. They proposed an algorithm based on multiple empirical kernel which maps data explicitly in multiple kernel spaces.

3 Methods

In this section, at first, we explain the SPM model which is used to represent images. Then, we introduce the designed feature fusion-based LMKL algorithm and its

optimization problem in detail. Finally, the optimization strategy to solve the problem is discussed.

3.1 Image representation by SPM model

Introducing the bag of word (BoW) model to compute image feature significantly improves the performance of image classification systems [30]. Pyramid matching is a BoW based model to approximate the similarity between two images [31]. In this model, a pyramid of grids is placed on the feature space at different resolutions. At each resolution level, the corresponding histogram of the image is computed. The weighted sum of histograms is computed such that finer resolutions get higher weights. Finally, the intersection kernel is applied on the weighted histograms of two images to approximate their correspondence. The main shortcoming of the pyramid matching method is that it discards the spatial information of images which plays an important role in the performance of image classification systems. Lazebnik et al. proposed the spatial pyramid match (SPM) approach to address the mentioned problem [1]. By extending BoW, the SPM method divides the original image into sub-regions in a pyramid manner and computes histograms of features in each sub-region separately. The final representation of the image is the concatenation of extracted histograms.

3.2 Preliminaries and formulation of feature fusion based LMKL

Consider the classification task as $D = \{(x_i, y_i)\}_{i=1}^N$ where N is the number of samples, x_i denotes the i^{th} sample and $y_i = \{\pm 1\}$ is the corresponding label for binary classification. In the MKL framework, multiple kernels are combined as follows:

$$K(x_i, x_j) = \sum_{k=1}^m \pi_k K_k(x_i, x_j) \quad (1)$$

where m is the number of kernels and π_k is the weight of k^{th} kernel.

The discriminator function $f(x_i)$ for a test data x_i in the standard MKL framework is formulated as follows:

$$f(x_i) = \sum_{k=1}^m \pi_k \langle w_k^T, \varphi_k(x_i) \rangle + b \quad (2)$$

where $\varphi_k(x_i)$ represents the k^{th} mapping function, and w_k and b are SVM parameters.

The standard framework of MKL assigns fixed weights to kernels in the entire space. As discussed in section 1.2, because of the large intraclass variance and inter class relationship in complicated spaces, such as an image feature space, similar weights for kernels are not suitable. For example, in some cases the kernel based on color

information is more informative than a texture based kernel. Therefore, a more accurate classifier will be achieved if variable weights are assigned to a kernel in different areas of the space.

Gönen and Alpaydin proposed a localized MKL (LMKL) framework in which the weights of kernels are calculated distinctly for each training sample [19]. The localized version of $K(x_i, x_j)$ is as follows:

$$K(x_i, x_j) = \sum_{k=1}^m \pi_k(x_i) \pi_k(x_j) K_k(x_i, x_j) \quad (3)$$

where $\pi_k(x_i)$ is the weight of k^{th} kernel corresponding to x_i .

In the original LMKL framework, Gönen et al. assumed that kernels are computed based on a single feature. In the proposed algorithm, multiple kernels are computed based on multiple features. Using multiple features instead of a single one results in a more accurate classifier in an image classification task as discussed in section 1.1. The kernel value between two images x_i and x_j is computed as follows:

$$K(x_i, x_j) = \sum_{k=1}^m \pi_k(x_i^k) \pi_k(x_j^k) K_k(x_i^k, x_j^k) \quad (4)$$

where x_i^k is a representation of training sample x_i corresponding to the k^{th} feature.

The combined kernel of (4) changes the standard kernel based margin maximization problem of SVM into a non-convex optimization problem. Instead of solving this difficult optimization problem, Gönen et al. estimated kernel weights by using the gating function.

A gating function formulates the effectiveness of the k^{th} kernel in classification of sample x_i . There are several ways to calculate the gating function. Sigmoid function formulated in (5) is a good choice and was used by Gönen et al. [19]:

$$\pi_k(x_i^k) = 1 / (1 + \exp(-\langle v_k, x_i^k \rangle - v_{k0})) \quad (5)$$

where v_k and v_{k0} are the parameters of the gating function. As stated before, x_i^k is a representation of training sample x_i corresponding to k^{th} feature which is in the form of a SPM histogram. Comparing the SPM histograms by their inner product is not accurate enough. X^2 kernel is a better choice in histogram comparison. Therefore, we modified the gating function of (5) by using the X^2 kernel instead of the inner product. The X^2 kernel based gating function is as follows:

$$\pi_k(x_i^k) = 1 / (1 + \exp(-X^2(v_k, x_i^k) - v_{k0})) \quad (6)$$

where X^2 kernel is defined as:

$$X^2(v_k(i), x(i)) = 2v_k(i)x(i) / (v_k(i) + x(i)), i = 1 \dots DG \quad (7)$$

DG is the dimension of feature space.

Because of the efficiency of X^2 kernel in computing the similarity of SPM histograms, we use the following gating function as well:

$$\pi_k(x^k) = X^2(v_k, x^k) + v_{k0} \quad (8)$$

3.3 Optimization strategy

By plugging local kernel weights in standard MKL formulation, the following optimization problem will result:

$$\begin{aligned} \min_{\{w_k\}, b, \{\xi_i\}, \{v_k\}, \{v_{k0}\}} & \frac{1}{2} \sum_{k=1}^m \|w_k\|^2 + C \sum_{i=1}^N \xi_i \quad (9) \\ \text{subject to} & \quad y_i \left(\sum_{k=1}^m \pi_k(x_i^k) \langle w_k, \Phi_k(x_i^k) \rangle + b \right) \geq 1 - \xi_i \\ & i = 1 \dots N, \quad \xi_i \geq 0 \end{aligned}$$

where C is the regularization parameter and ξ_i s are the slack variables.

Since standard MKL is a convex optimization problem it can be solved by common optimization methods. Combining nonlinear gating functions with standard MKL problem changes the convex optimization problem of MKL into a nonlinear and non-convex problem. This problem can be solved using the alternate optimization method, which is an iterative two step approach. In step one, some parameters are assumed to be fixed and the others are computed by solving the optimization problem. In step two, the non-fixed parameters in the first step are considered to be fixed and the remaining parameters are calculated by solving the new optimization problem. The optimization algorithm iterates until convergence. We considered two termination criteria: the maximum number of iterations and reaching the changes of object function below a predefined threshold.

Step one: Learning SVM parameters.

In this step, the optimization problem should be minimized with respect to w_k , ξ_i and b , while v_k and v_{k0} are fixed. In order to remove the constraints, the Lagrangian of problem (9) is calculated and the following problem is obtained:

$$\begin{aligned} L(\{w_k\}, b, \{\xi_i\}, \{\lambda_i\}, \{\eta_i\}) &= \frac{1}{2} \sum_{k=1}^m \|w_k\|^2 + \sum_{i=1}^N (C - \lambda_i - \eta_i) \xi_i \\ &+ \sum_{i=1}^N \lambda_i - \sum_{i=1}^N \lambda_i y_i \left(\sum_{k=1}^m \pi_k(x_i^k) \langle w_k, \Phi_k(x_i^k) \rangle + b \right) \end{aligned} \quad (10)$$

where λ_i and η_i are Lagrangian parameters.

Calculating the derivatives of (10) with respect to $\{w_k\}$, b and ξ_i will result in:

$$\begin{aligned} \partial L / \partial w_k &= 0 \Rightarrow w_k - \sum_{i=1}^N \lambda_i y_i \pi_k(x_i^k) \Phi_k(x_i^k) = 0 \\ \partial L / \partial b &= 0 \Rightarrow \sum_{i=1}^N \lambda_i y_i = 0 \\ \partial L / \partial \xi_i &= 0 \Rightarrow C - \lambda_i - \eta_i = 0 \end{aligned} \quad (11)$$

substituting (11) in (10), the dual problem of (10) is obtained:

$$\begin{aligned} J &= \max_{\{\lambda_i\}} \sum_{i=1}^N \lambda_i - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j y_i y_j \sum_{k=1}^m \pi_k(x_i^k) \pi_k(x_j^k) K_k(x_i^k, x_j^k) \\ \text{such that} & \quad \sum_{i=1}^N \lambda_i y_i = 0, \quad 0 \leq \lambda_i \leq C \end{aligned} \quad (12)$$

where $K_k(x_i^k, x_j^k) = \Phi_k(x_i^k) \Phi_k(x_j^k)$.

If we prove that the localized weighted sum of kernels $\sum_{k=1}^m \pi_k(x_i^k) \pi_k(x_j^k) K_k(x_i^k, x_j^k)$ is a positive semi definite kernel matrix, then (12) can be solved as a standard canonical SVM problem.

In order to prove that the localized weighted sum of kernels is positive semi definite, we use the definition of a quasi-conformal transformation. For a positive function $c(x)$, a quasi-conformal transformation of $K(x, y)$ is defined as follows:

$$\tilde{K}(x, y) = c(x)c(y)K(x, y) \quad (13)$$

The gating function in (6) and (8) used in our experiments always provide positive values; therefore, $\pi_k(x_i^k)$ $\pi_k(x_j^k) K_k(x_i^k, x_j^k)$ in (4) is a quasi-conformal transformation of $K(x, y)$. Positive semidefinite kernels are closed under quasi-conformal transformation [32], so $\pi_k(x_i^k) \pi_k(x_j^k) K_k(x_i^k, x_j^k)$ is a positive semi-definite kernel. On the other hand, summing up several kernels together leads to a single kernel. Thus, $\sum_{k=1}^m \pi_k(x_i^k) \pi_k(x_j^k) K_k(x_i^k, x_j^k)$ is a positive semidefinite kernel as well and (12) is considered as a canonical SVM that can be solved by common approaches.

Step Two: Learning locality function parameters.

To determine the values of parameters in gating functions, we use the gradient descent method such that the derivatives of the dual problem of (12) are calculated with respect to ν_k and ν_{k0} while $\{w_k\}$, b and ξ_i are fixed. The step size of each iteration is determined by a line search method. Taking derivatives of problem (12) with respect to ν_k and ν_{k0} we obtain:

$$\begin{aligned} \partial J / \partial \nu_k &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^m \lambda_i \lambda_j y_i y_j K_k(x_i^k, x_j^k) (\pi_k'(x_i^k) \pi_k(x_j^k) \\ &\quad + \pi_k(x_i^k) \pi_k'(x_j^k)) \\ \partial J / \partial \nu_{k0} &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^m \lambda_i \lambda_j y_i y_j \pi_k(x_i^k) \pi_k(x_j^k) \\ &\quad K_k(x_i^k, x_j^k) (2 - \pi_k(x_i^k) - \pi_k(x_j^k)) \end{aligned} \quad (14)$$

where $\pi_k'(x)$ is defined as (15) for the X^2 gating function of (8),

$$\{2(x(i)(x(i) + \nu_k(i)) - x(i)\nu_k(i)) / (x(i) + \nu_k(i))^2\}_{i=1}^{DG} \quad (15)$$

also $\pi_k(x)$ is defined as (16) for the X^2 kernel based sigmoid function of (6),

$$A(\exp(-X^2(\nu_k, x^k) - \nu_{k0})) / (1 + \exp((-X^2(\nu_k, x^k) - \nu_{k0})))^2 \quad (16)$$

where A is equal to (15).

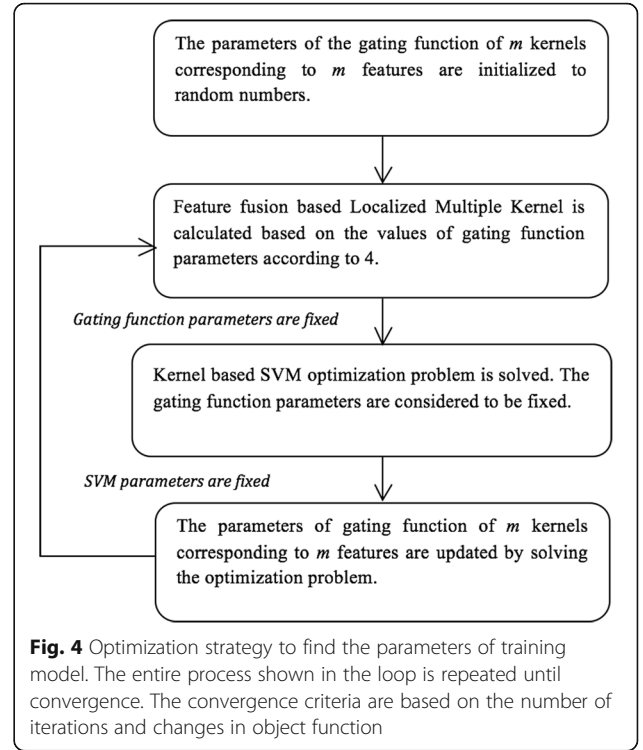
The block diagram of the optimization strategy to find the parameters of the training model is depicted in Fig. 4.

4 Results and discussion

In this section, we conduct some experiments to study the classification performance of the proposed method on two widely used benchmark datasets: Caltech 101 [33] and Caltech 256 [34]. The mentioned datasets are challenging for image classification because of their large intra class variance and inter class relationship. In particular, in Caltech 256 the intra class variance is very large, making it more challenging for image classification.

4.1 Experimental configurations

We explain the implementation details of our proposed algorithm in this section. To describe images, at first, the features are extracted, then the kernels are computed based on them. We used the subset of features suggested in [15]. The selected features for Caltech 101 include dense SIFT (scale invariant feature transform) [30], dense



color SIFT and SSIM (structural similarity) [35]. Dense SIFT is calculated over regular grids of 16×16 image patches with eight pixels spacing using VLFeat Lib [36]. Likewise, color-dense SIFT is calculated in three channels of CIElab. SSIM is computed in 5×5 patches to obtain a correlation map.

To represent images for classification, we considered spatial pyramid match (SPM) histograms based on the extracted features [1]. To this end, we trained three separate dictionaries via k-means clustering for dense SIFT, color dense SIFT and SSIM feature spaces. The numbers of visual words for each individual dictionary are 600, 600, and 300, respectively. Compared to similar works, we used less visual words for each dictionary,

Table 1 Performance comparison of algorithms on Caltech 101 using 30 training images per class

Algorithms	Accuracy in %
SVM-KNN [37]	66.20
SPM [1]	64.40
ScSPM [38]	73.20
NBNN [39]	70.40
LLC [40]	73.44
BT-Gist [42]	63.8
Structured LR [43]	73.6
Wang et al. [2]	72.8
Proposed method	81.92

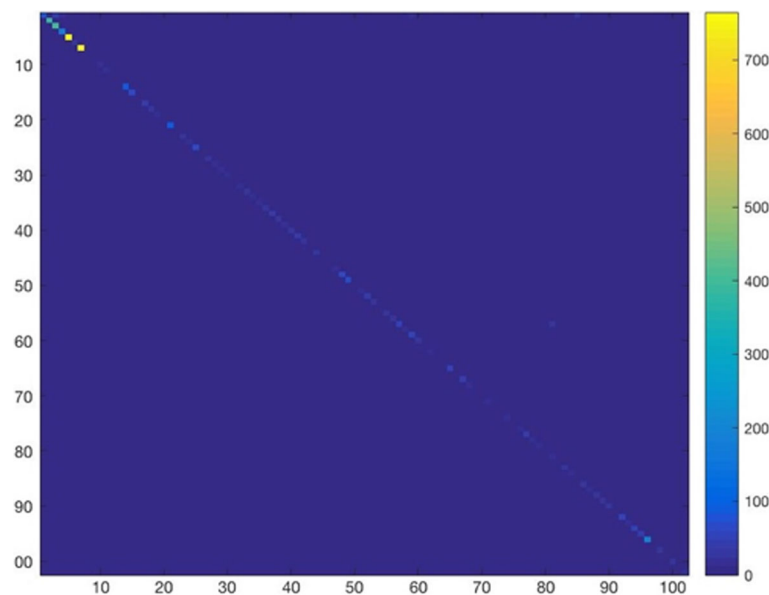


Fig. 5 Confusion matrix of Caltech 101 classification by the proposed algorithm

thereby avoiding large feature vectors. As a result, the computation time is reduced. To generate the SPM representation, each image was partitioned hierarchically into 1×1 , 2×2 and 4×4 blocks and the corresponding feature vectors of each individual block was encoded based on the learned dictionaries.

The abovementioned SPM based feature vectors were fed to the proposed classifier. To compute the train-train and train-test kernel matrices, we used the parameter free X^2 kernel for all features. The proposed algorithm is written in MATLAB and the source codes available in [15, 23] are used as well.

We used two gating functions to compute the kernel weights: X^2 based sigmoid and X^2 as formulated in (6) and (8). We partitioned the training data to train set and validation set by cross validation. Then, we grid searched the space to tune the SVM regularization parameter and the gating function simultaneously. The SVM regularization

parameter is set to 10 and X^2 is selected as the gating function by cross validation.

The optimization problem discussed in section 2, was solved in two phases in an iterative manner. In the first phase, the parameters of gating function are fixed and the problem is solved in the same method as a standard kernel based SVM problem. In the second phase, the problem is solved to find the parameters of the gating function by a gradient descent approach.

In addition, we followed the One vs. All strategy in the training phase where we trained one classifier for each individual class. We should note that, generally compared to the One vs. One method, the One vs. All method suffers from high data imbalance between one class and the remaining classes. However, because of the high intraclass variance in real world image classification, the One vs. One method suffers from the same high data imbalance problem. The data imbalances both inside each class and between classes are addressed by dedicating variable weights to kernels as discussed in section 1.2.

Table 2 Performance comparison of algorithms on Caltech 256 using 30 training images per class

Algorithms	Accuracy in %
SPM [1]	34.1
ScSPM [38]	34.02
NBNN [39]	37.0
LLC [40]	41.19
Oliveira et al. [44]	37.08
Wang et al. [2]	45.2
Proposed method	42.12

4.2 Evaluations on Caltech 101

Caltech 101 contains a total of 9144 images in 101 object classes and an extra BACKGROUND class [33]. Each class has 31 to 800 images. The size of most images is medium, about 300×300 . Caltech 101 is a challenging dataset because of the large number of classes, intra class variance, and interclass relationship. For fair comparison with other works, we followed the experimental setup suggested in [1] and randomly selected 30 images per class for training, leaving the rest for testing.

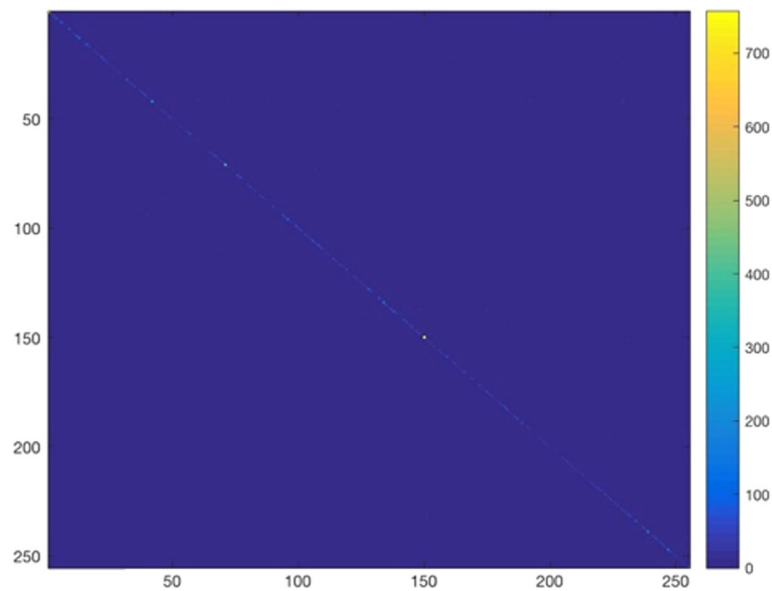


Fig. 6 Confusion matrix of Caltech 256 classification by the proposed algorithm

Table 1 reports the mean classification accuracy over 102 classes in Caltech 101. It shows the reported performance of the related algorithms and ours. According to this table, our algorithm outperforms all of the baseline algorithms including nearest neighbor-based SVM [37], SPM [1], ScSPM [38], nearest neighbor [39], and LLC [40].

In addition, we note that as reported in [15], which has the same experimental setup as ours, the classification accuracy using single kernel based SVM is around

73, 62.5, and 62% for dense SIFT, color dense SIFT, and SSIM features, respectively. The confusion matrix of the classification is depicted in Fig. 5.

4.3 Evaluations on Caltech 256

Caltech 256 contains 30,607 images in 256 classes and a BACKGROUND class [34]. Each class contains at least 80 images. Compared to Caltech 101, Caltech 256 is more

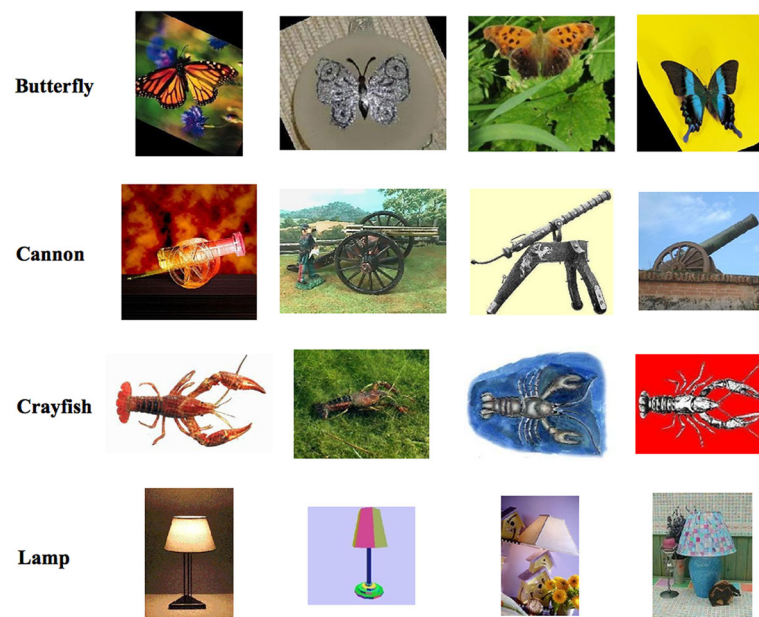


Fig. 7 A few instance images from some difficult classes from Caltech 101. This figure illustrates large intra class variance in each class

challenging because the objects are not centered in the images and the intra class variance is much higher.

As a common experimental setup for this dataset, we chose 30 images per class for training and used the rest for testing. We measured the performance of our proposed algorithm by calculating the mean classification accuracy over 257 classes. Table 2 shows the comparison results of our algorithm with the related ones. Fig. 6 illustrates the classification confusion matrix.

As seen in Table 2, the classification accuracy of [2] is 3.08% better than ours. The reason for this better performance is that, in comparison to SPM (the feature extraction used in our algorithm), the method in [2] not only considers the spatial information of images, but also the shape information. To this end, they integrate the salient region and the spatial geometry structure. This combination makes the visual words more discriminative. In addition, this integration makes the extracted feature vectors more resistant to both the complexity of background and location variations of images in each category. This approach indirectly gives more weight to shape descriptor parameters which could be the cause of better performance of this method on large datasets.

5 Performance on difficult classes

There are some classes in Caltech 101 in which images are very difficult to be classified because of the high intra class variance. In [41] the average classification accuracy for nine difficult classes including butterfly, crab, cannon, crayfish, beaver, crocodile, cougar body, chair and lamp, is reported as 24%, while our proposed method has an average accuracy of 52.38% for the same classes. Fig. 7, shows samples from four of these difficult classes.

In addition, [1] has tested their method on four difficult classes which are cougar body, beaver, crocodile and ant from Caltech 101 and reported the classification accuracy for each individual class. We compared the performance of our method with [1] on the same classes. The results are as shown in Table 3.

We should note that in our proposed method, the improvement of classification accuracy on difficult classes is the result of calculating the local weights for kernels which could address the problem of high intra class variance.

Table 3 Comparison of the proposed method with [1] on individual difficult classes on Caltech 101

	Classification accuracy in % for difficult classes			
	Cougar body	Beaver	Crocodile	Ant
SPM [1]	27.6	27.5	25.0	25.0
Proposed method	47.06	18.75	50.0	25.0

6 Conclusions

Image classification, which is the task of determining the semantic class of un-labeled test samples, is a challenging task especially for real world images. Two issues challenge the classification accuracy in image classification. First, images are better described by several types of features; thus, the designed system should be able to merge heterogeneous features. The second challenge comes from the large intraclass variance and interclass relationship in real world image databases.

In this study, we designed a feature fusion-based localized multiple kernel learning algorithm using the SPM feature to overcome the mentioned difficulties. Our results demonstrate that the proposed approach performs well in image classification problems. The higher performance of our method partially depends on computing weights of kernels locally. In the future, we will directly compute kernel weights in the kernel space.

Acknowledgements

Not applicable.

Availability of data and materials

Not applicable.

Authors' contributions

Both authors designed the proposed algorithm together. FZ implemented it with MATLAB. Both authors read and approved the final manuscript.

Funding

We would like to thank Iran Telecommunication Research Centre for their support of this research.

Authors' information

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 March 2017 Accepted: 6 November 2017

Published online: 29 November 2017

References

1. S Lazebnik, C Schmid, J Ponce, Beyond BAGs of Features Spatial Pyramid Matching for Recognizing Natural Scene Categories. Paper Presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006
2. R Wang, K Ding, J Yang, A novel method for image classification based on bag of visual words. *J. Vis. Commun. Image Represent.* **40**, 24–33 (2016)
3. P Zheng et al., Image set classification based on cooperative sparse representation. *Pattern Recogn.* **63**, 206–217 (2017)
4. M Yang, H Chang, W Luo, Discriminative analysis-synthesis dictionary learning for image classification. *Neurocomputing* **219**, 404–411 (2017)
5. V Abrol, P Sharma, A Sao, Greedy dictionary learning for kernel sparse representation based classifier. *Pattern Recogn. Lett.* **78**, 64–69 (2016)
6. X Yuan, X Liu, S Yan, Visual classification with multitask joint sparse representation. *IEEE Trans. Image Process.* **21**, 4349–4360 (2012)
7. A Shrivastava, V Patel, R Chellappa, Multiple kernel learning for sparse representation-based classification. *IEEE Trans. Image Process.* **23**, 3013–3024 (2014)
8. S Zhang et al., Constructing deep sparse coding network for image classification. *Pattern Recogn.* **64**, 130–140 (2017)

9. S Ding, L Guo, Y Hou, Extreme learning machine with kernel model based on deep learning. *Neural Comput. & Applic.* **28**, 1975–1984 (2016).
10. M Uzair, F Shafait, B Ghanem, A Mian, Representation learning with deep extreme learning machines for efficient image set classification. *Neural Comput. & Applic.* 1–13 (2015).
11. L Xie et al., Incorporating visual adjectives for image classification. *Neurocomputing* **182**, 48–55 (2016)
12. Y Yeh et al., A novel multiple kernel learning framework for heterogeneous feature fusion and variable selection. *IEEE Trans Multimedia* **14**, 563–574 (2012)
13. H Wang, G Fu, Y Cai, S Wang, Multiple Feature Fusion Based Image Classification Using a Non-biased Multi-Scale Kernel Machine. Paper Presented at the 12th International Conference on Fuzzy Systems and Knowledge Discovery, Zhangjiajie, China, 15–17 August 2015
14. B Fernando, E Fromont, D Muselet, M Sebban, Discriminative Feature Fusion for Image Classification. Paper Presented at the 12th IEEE Conference on Computer Vision and Pattern Recognition, Providence, Rhode Island, 16–21 June 2012
15. A Vedaldi, M Varma, V Gulshan, A Zisserman, VGG - Multiple Kernels for Image Classification. <http://www.robots.ox.ac.uk/~vgg/software/MKL>. Accessed 21 Mar 2017.
16. S Shafiee, F Kamangar, V Athitsos, J Huang, L Ghandehari, Multimodal Sparse Representation Classification with Fisher Discriminative Sample Reduction. Paper Presented at IEEE International Conference on Image Processing, Paris, France, 27–30 October 2014
17. J Shawe-Taylor, N Cristianini, *Kernel Methods for Pattern Analysis*. (Cambridge, Cambridge University Press, 2004).
18. G Lanckriet, N Cristianini, P Bartlett, L El Ghaoui, M Jordan, Learning the kernel matrix with semidefinite programming. *J Mach Learn Res* **5**, 27–72 (2004)
19. M Gönen, E Alpaydin, Localized Multiple Kernel Learning. Paper Presented in Proceedings of the 25th International ACM Conference on Machine Learning, New York, NY, USA, 05–09 July, 2008
20. Y Gu, Q Wang, X Jia, JA Benediktsson, A novel MKL model of integrating LiDAR data and MSI for urban area classification. *IEEE Trans. Geosci. Remote Sens.* **10**, 5312–5326 (2015)
21. Q Wang, Y Gu, D Tuia, Discriminative multiple kernel learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **54**, 3912–3927 (2016)
22. Y Member, T Liu, X Jia, JA Benediktsson, J Chanussot, Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **54**, 3235–3247 (2016)
23. M Gönen, E Alpaydn, Multiple kernel learning algorithms. *The. J. Mach. Learn. Res.* **12**, 2211–2268 (2011)
24. D Lewis, T Jebara, W Noble, Nonstationary Kernel Combination. Paper presented at the 23rd international conference on Machine learning, Pittsburgh, Pennsylvania, USA, 25–29 June 2006
25. W Lee, S Verzakov, R Duin, Kernel Combination Versus Classifier Combination. Multiple Classifier Systems, Paper Presented at the 7th International Workshop on Multi Classifier Systems, Prague, Czech Republic, Springer, 23–25 May 2007
26. J Yang, Y Li, Y Tian, L Duan, W Gao, Group-Sensitive Multiple Kernel Learning for Object Categorization. Paper Presented at the IEEE International Conference on Computer Vision, Kyoto, Japan, 29 September - 2 October 2009
27. R Kannao, P Guha, Success based locally weighted multiple kernel combination. *Pattern Recogn.* **68**, 38–51 (2017)
28. J Lu, G Wang, P Moulin, Image Set Classification Using Holistic Multiple Order Statistics Features and Localized Multikernel Metric Learning. Paper Presented at the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013
29. Q Fan, D Gao, Z Wang, Multiple empirical kernel learning with locality preserving constraint. *Knowl.-Based Syst.* **105**, 107–118 (2016)
30. D Lowe, Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**, 91–110 (2004)
31. K Grauman, T Darrell, The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features. Paper presented at the IEEE International Conference on Computer Vision, Beijing, China, 15–21 October 2005
32. S Amari, S Wu, Improving support vector machine classifiers by modifying kernel functions. *Neural Netw.* **12**, 783–789 (1999)
33. L Fei-Fei, R Fergus, P Perona, Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Comput. Vis. Image Underst.* **106**, 59–70 (2007)
34. G Griffin, A Holub, P Perona, Caltech-256 Object Category Dataset. <http://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001>. Accessed 21 Mar 2017.
35. E Shechtman, M Irani, Matching Local Self-Similarities across Images and Videos. Paper Presented at the IEEE International on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007
36. A Vedaldi, B Fulkerson, VLFeat: An Open and Portable Library of Computer Vision Algorithms. Paper Presented at the 18th ACM International Conference on Multimedia, Firenze, Italy, 25–29 October 2010
37. H Zhang, SVM-KNN: Discriminative Nearest Neighbor Classification for Visual Category Recognition. Paper Presented at the IEEE International Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006
38. J Yang, K Yu, Y Gong, T Huang, Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. Paper Presented at the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009
39. O Boiman, E Shechtman, M Irani. In Defense of Nearest-Neighbor Based Image Classification. Paper Presented at the IEEE International Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, USA, 24–26 June 2008
40. J Wang, J Yang, K Yu, F Lv, T Huang, Y Gong, Locality-Constrained Linear Coding for Image Classification. Paper Presented at the IEEE International Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 13–18 June 2010
41. K Hotta, Object Categorization Based on Kernel Principal Component Analysis of Visual Words. Paper Presented at the IEEE Workshop on Applications of Computer Vision, Copper Mountain, Colorado, 7–9 Jan 2008
42. Y Han, G Liu, Biologically inspired task oriented gist model for scene classification. *Comput. Vis. Image Underst.* **117**, 76–95 (2013)
43. Y Zhang, Z Jiang, L Davis, Learning Structured LowRank Representations for Image Classification, Paper Presented at the IEEE International Conference on Computer Vision and Pattern Recognition, Portland, Oregon, 25–27 June 2013
44. GL Oliveira, ER Nascimento, AW Vieira, Sparse spatial coding: A novel approach for efficient and accurate object recognition. International Conference on Robotics and Automation, St. Paul, MN, USA, 14–18 May 2012

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com